Applying genetic epidemiology to explore sex-specific causes, correlations and consequences of obesity



Jenny Charlotta Censin

Nuffield Department of Medicine
University of Oxford
Lincoln College, Hilary Term 2021

Thesis submitted to the University of Oxford for the degree Doctor of Philosophy

Memorandum

The work in this thesis is a result of my own original work, except as declared in this memorandum or specified in the text.

The data resources that have been used in the analyses have all undergone various degrees of central processing before inclusion in my data analysis pipelines. Throughout my thesis, I have clearly described which quality control measures have been conducted by others and which procedures that have been performed by me. A few of the scripts used to generate my results, such as those used for extracting diabetes cases, are from previous publications. In such cases, this is clearly explained and referenced in the text.

I have also conducted all the analyses and generated all the tables and figures in the publications arising from the work included in this thesis, *except* for the article pertaining to miscarriage (included as a part of Chapter 4). My contribution to the article about miscarriage, by Laisk *et al.* (2020), was restricted to the colocalisation analyses. I had no part in conducting the genome-wide association studies or MRs in the article, but I conducted the MRs included in Chapter 4 separately (Laisk *et al.*, 2020).

This thesis is approximately 49,850 words.

Acknowledgements

I would first like to thank my supervisors; my primary supervisors Professor Cecilia Lindgren and Professor Michael Holmes, and with the added support of Professor Mark McCarthy and informal supervision and mentoring from Professor Tove Fall. You have taught me to think critically, independently, and how to overcome seemingly impossible problems. Under your guidance, I have grown both as a researcher and as a person. Thank you.

I would also like to express my gratitude to my friends and colleagues who were there to discuss scientific questions and to offer support; Dr. Jonas Bovijn, Dr. Teresa Ferreira, Dr. Sara Pulit, Dr. Laura Wittemans, Dr. Triin Laisk, and Professor Christian Berne. I would also like to express a special thank you to Stefanía, Helen, Rosemary, and Peter – your friendship, kindness, and support were all invaluable to me.

To my friends and family – I love you all. To my family, thank you for always being there for me; Daniel, Christian, Madde, Siv, Knut, Trassel, and Lucifer. To my closest friends, thank you for offering a sympathetic ear when needed; Alexandra, Ellen, Viktoria, Lisa, and Linn. Finally; Nick, thank you for always having my back. Pappa; thank you for always being there for a chat. Mamma; thank you for your relentless support when things were hard. I could not have done it without you.

Table of Contents

Memorandum ____ *i*

Acknowledgements ____ *ii*

Table of Contents ____ *iii*

Abstract ____ *ix*

Abbreviations ____ *xi*

Chapter 1. Introduction ____ *1*

1.1. Sex matters! ____ **1**

1.1.1. A call for including sex as a biological variable in research ____ 1

1.2. The obesity pandemic – a problem for both men and women ____ **2**

1.2.1. The rise in obesity rates ____ 2

1.2.2. Measures of obesity ____ 3

1.2.3. A short background to pathophysiological mechanisms of obesity ____ 4

1.2.4. The global burden of obesity ____ 5

1.2.5. The current state of drugs and biomarkers in obesity-related disorders ____ 9

1.3. Obesity and genetics ____ **14**

1.3.1. An introduction to genome-wide association studies ____ 14

1.3.2. The role of genetics in obesity ____ 17

1.4. Using genetic epidemiology to tease out relationships between traits ____ **19**

1.4.1. Leveraging genetic associations to explore associations between traits ____ 19

1.4.2. Mendelian randomisation ____ 20

1.4.3. Colocalisation ____ 30

1.5. Thesis aims ____ **34**

Chapter 2. Sex-specific effects of obesity on the leading causes of death ____ *37*

2.1. Introduction 37

2.1.1. Previous studies of sex-specific effects of obesity on leading causes of death have been inconclusive 37

2.1.2. The importance of teasing out sex-specific consequences of obesity 39

2.1.4. Aims 40

2.2. Methods 41

2.2.1. Study overview 41

2.2.2. The UK biobank 44

2.2.3. Obesity traits 46

2.2.4. Sex-specific genetic risk scores for obesity 47

2.2.7. Statistical analyses 54

2.2.8. Sensitivity analyses 56

2.2.9. Data availability and software used 59

2.3. Results 59

2.3.1. Participant characteristics 59

2.3.2. Instrument strength 61

2.3.3. Sex-specific effects of obesity 64

2.3.4. Sensitivity analyses 68

2.4. Discussion and Conclusion 78

2.4.1. Instruments 78

2.4.2. Diabetes and glycaemic traits 79

2.4.3. Cardiovascular disease 80

2.4.4. Lung disease and smoking 81

2.4.5. Renal failure 82

2.4.6. Liver disease 83

2.4.7. Other diseases 83

2.4.8. Mendelian randomisation and sensitivity analyses 84

2.4.9. Conclusion 88

Chapter 3. Sex-specific effects of proteins on obesity and cardiometabolic disease ____ *90*

3.1. Introduction ____ **90**

3.1.1. The promise of proteomics ____ 90

3.1.2. The absence of sex-stratified genetic association studies of proteins ____ 91

3.1.3. Aims ____ 91

3.2. Methods ____ **92**

3.2.1. Study overview ____ 92

3.2.2. Datasets ____ 93

3.2.3. Observational correlations between pairs of aptamers ____ 99

3.2.4. Post-hoc functional enrichment analyses ____ 100

3.2.5. Sex-specific genome-wide association studies of protein levels ____ 101

3.2.6. Sex-specific genetic aptamer-aptamer correlations ____ 102

3.2.7. Sex-stratified genome-wide association studies of type 2 diabetes and coronary artery disease ____ 102

3.2.8. Mendelian randomisation analyses of proteins on cardiometabolic traits ____ 103

3.2.9. Sensitivity analyses ____ 108

3.2.10. Data availability and software ____ 113

3.3. Results ____ **114**

3.3.1. Wide-spread correlations between aptamers using observational data ____ 114

3.3.2. Functional enrichment analysis highlights hormones and reproductive traits ____ 116

3.3.3. Sex-stratified genome-wide association studies of protein levels detect novel, sex-specific effects ____ 121

3.3.4. Genetic aptamer-aptamer correlation analyses ____ 122

3.3.5. Sex-stratified Mendelian randomisation analyses identify sex-specific effects of proteins on obesity ____ 123

3.3.6. Sensitivity analyses ____ 125

3.4. Discussion and Conclusion ____ **131**

3.4.1. Protein-protein correlations in the human plasma proteome 131
3.4.2. Sex-stratified genome-wide association studies of protein levels 133
3.4.3. Mendelian randomisation analyses 134
3.4.4. Limitations inherent to the data 138
3.4.5. Sensitivity analyses 142
3.4.6. Conclusion 143

Chapter 4. Using colocalisation to identify potential mediators in obesity-related female reproductive diseases 144

4.1. Introduction 144

4.1.1. The role of obesity in female reproductive diseases 144
4.1.2. The genetics of polycystic ovary syndrome and miscarriage 145
4.1.3. Female reproductive diseases – common and with far-reaching consequences, but often overlooked 146
4.1.4. Aims 147

4.2. Methods 147

4.2.1. Study overview 147
4.2.2. Datasets 148
4.2.3. Mendelian randomisation analyses 155
4.2.4. Primary evaluation of eCAVIAR 157
4.2.5. Main colocalisation analyses performed using Coloc 158
4.2.6. Sensitivity analyses 161
4.2.7. Phenome-wide association study and in-silico investigations 165
4.2.8. Data availability and software 165

4.3. Results 166

4.3.1. Mendelian randomisation analysis confirms a role of obesity in polycystic ovary syndrome, but not in miscarriage 166
4.3.2. eCAVIAR results are sensitive to choice of settings 167

4.3.3. Inadvertently selecting genetic variants based on P-value may yield false positive results 171
4.3.4. Main Coloc results 173
4.3.5. Sensitivity analyses 178
4.3.6. Regulatory functions and associations with other traits 183
4.4. Discussion and Conclusion 183
4.4.1. Mendelian randomisation analyses 184
4.4.2. Evaluation of eCAVIAR 184
4.4.3. Selecting which genetic variants to include based on P-value may bias results 185
4.4.4. Colocalisation analyses of polycystic ovary syndrome 186
4.4.5. Conclusion 191
Chapter 5. Discussion and Conclusion 192
5.1. Discussion 192
5.1.1. A summary 192
5.1.2. Key limitations 194
5.1.3. Directions for future research 198
5.2. Conclusion 199
References 201
Appendix 252
Appendix 1. Data acknowledgements, funding, publications, and data availability 252
Appendix 1.1. Data acknowledgements 252
Appendix 1.2. Funding 253
Appendix 1.3. Publications 254
Appendix 1.4. Data availability 256
Appendix 2. Supplement to Chapter 2 258
Appendix 3. Supplement to Chapter 3 304

Appendix 4. Supplement to Chapter 4 _____ 405

Abstract

Obesity has considerable consequences on global health. Yet, it is relatively unexplored if the causes and consequences of obesity and obesity-related diseases differ between men and women. The increased scope of genetic studies provides an opportunity to leverage genetic data to characterise sex-specific epidemiological relationships using methods less prone to reverse causation and confounding, such as Mendelian randomisation. Thus, I aimed to investigate the sex-specific causes and consequences of obesity using genetic epidemiology, as well as to identify potential disease-mediating genes in two common obesity-related female reproductive diseases: polycystic ovary syndrome and miscarriage.

In my first research chapter, I use Mendelian randomisation to investigate the sex-specific effects of obesity on leading causes of death from non-communicable diseases, before comparing the men- and women-specific effect estimates. My results indicate that the disease risk arising from obesity differs between the sexes for several diseases, including type 2 diabetes, renal failure, and chronic obstructive pulmonary disease.

I then proceed to perform sex-stratified genome-wide association studies of 2,623 unique proteins and compare the effect estimates of the genetic variants between the sexes. Using these data, I compare the effects of proteins on obesity and cardiometabolic diseases between men and women. My results in this chapter indicate that up to 15% of protein-associated genetic variants may have different effects in the two sexes. I further identify seven proteins with sexually heterogeneous effects on obesity in men versus women, including two proteins – kynureninase and N-terminal pro b-type natriuretic peptide – with seemingly opposite effects in men compared to women. However, sensitivity analyses indicated that my results may be affected by poor specificity in the original protein measurements.

Finally, in my third research chapter, I use colocalisation to highlight potential disease-mediating genes in polycystic ovary syndrome and recurrent miscarriage. While I was not able to identify any potential disease-mediating genes for recurrent miscarriage, I identify eleven genes and proteins with evidence of colocalising with risk of polycystic ovary syndrome.

My thesis demonstrates that causal relationships in cardiometabolic disease can differ between the sexes, and emphasises the importance of investigating sex-specific effects. Furthermore, my results highlight several genes and proteins that may contribute to obesity and obesity-related diseases. Although their clinical utility will need to be established in future studies, several of these genes and proteins hold promise for use as biomarkers or for targeting by drugs to improve the health of both men and women.

Abbreviations

B3GA3	Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3
BMI	Body mass index
CAD	Coronary artery disease
CAUSE	Causal Analysis Using Summary Effect estimates
CI	Confidence interval
CLD	Chronic liver disease
CLPP	Colocalisation posterior probability (from eCAVIAR)
COPD	Chronic obstructive pulmonary disease
DBP	Diastolic blood pressure
EA	Effect allele
eQTL	Expression quantitative trait locus
FDR	False discovery rate
FG	Fasting glucose
FI	Fasting insulin
FSH	Follicle-stimulating hormone
GCTA	Genome-wide Complex Trait Analysis
GIANT	Genetic Investigation of ANthropometric Traits
GRS	Genetic risk score (based on genetic variants that surpass a set multiple testing threshold)
GTEx	The Genotype-Tissue Expression project
GWAS	Genome-wide association study
h_g^2	SNP-based heritability
HWE	Hardy-Weinberg equilibrium
IVW	Inverse-variance weighted
KING	Kinship-based INference for Genome-wide association studies

LD	Linkage disequilibrium
LH	Luteinizing hormone
MR	Mendelian randomisation
N-terminal pro-BNP	N-terminal pro b-type natriuretic peptide
NAFLD	Non-alcoholic fatty liver disease
NCBI	National Center for Biotechnology Information
NEA	Non-effect allele
NMDA	N-methyl-D-aspartate
OR	Odds ratio
PC	Principal component
PCOS	Polycystic ovary syndrome
PCSK1	Neuroendocrine convertase 1
$P_{enrichment}$	P-value for the g:OST enrichment analysis in Chapter 3
P_{GWAS}	P-value for the genome-wide association studies in Chapter 3
P_{het}	P-value for Cochran's Q test for heterogeneity
P_{LDSC}	P-value for genetic correlations in Chapter 3
PP	Posterior probability
pQTL	Protein quantitative trait locus
$P_{sexdiff}$	P-value for a sex difference
$P_{correlation}$	P-value for correlations in Chapter 3
$P_{correlation\text{-}sexdiff}$	P-value for a sex difference in the correlations in Chapter 3
PoPS	Polygenic Priority Score
PSA	Prostate specific antigen
PTGD2	Hematopoietic prostaglandin D synthase
RFU	Relative fluorescent units
RINT	Rank-inverse normal transformation
SAIGE	Scalable and Accurate Implementation of GEneralized mixed model
SBP	Systolic blood pressure

SNP	Single-nucleotide polymorphism
T1D	Type 1 diabetes
T2D	Type 2 diabetes
UKBB	UK biobank
WHR	Waist-hip ratio
WHRadjBMI	Waist-hip ratio adjusted for body mass index

Chapter 1. Introduction

1.1. Sex matters!

1.1.1. A call for including sex as a biological variable in research

Men and women are not biologically identical. Sex differences are obvious in certain examples of body morphology, hormone and biomarker levels, and clinical presentations of diseases (Schreiner *et al.*, 1996; Demerath *et al.*, 2007; DeVon *et al.*, 2008; Karakas *et al.*, 2018; Khan *et al.*, 2018). There are numerous cases where such sex differences have a verified or potential impact on clinical care. For example, acute myocardial infarction present differently in men versus women (DeVon *et al.*, 2008), eight out of ten drugs withdrawn from the market in 1997-2001 had greater health risks for women than men (United States General Accounting Office, 2001a), and testosterone may *increase* type 2 diabetes (T2D) risk in women but *decrease* T2D risk in men (Karakas *et al.*, 2018; Ruth *et al.*, 2020). Yet, sex-specific effects have been – and are – frequently understudied, underreported, and/or unreleased, including for animal studies, drug trials, and genome-wide association studies (GWASs) (Meinert *et al.*, 2000; Beery and Zucker, 2011; Geller *et al.*, 2011, 2018; CARDIoGRAMplusC4D Consortium, 2020; DIAGRAM, 2020).

Historically, research was often undertaken in men only (Meinert *et al.*, 2000; United States General Accounting Office, 2001b; Geller *et al.*, 2011, 2018; Turcot *et al.*, 2018). In the 1970's and 80's, the American Food and Drug Administration recommended the exclusion of women of reproductive age from clinical trials (Food and Drug Administration, 1977; Liu and Mager, 2016). While the recommendation was largely driven by concerns of teratogenicity, a major consequence was a general underrepresentation of women in clinical trials (The Steering Committee of the Physicians' Health Study Research, 1988; Food and Drug Administration, 1993; Meinert *et al.*, 2000; United States General Accounting Office, 2001b; Liu and Mager, 2016). In 1993, the Food

and Drug Administration conceded that the policy *"perhaps delayed appreciation of gender-related variation in drug effects"* (Food and Drug Administration, 1993) and reversed their former recommendation. After this policy change, women's participation rates in clinical trials increased, but a majority (72%) of recent trials funded by the National Institutes of Health did still not report sex-specific outcomes (Meinert *et al.*, 2000; United States General Accounting Office, 2001b; Geller *et al.*, 2011, 2018).

The historical exclusion of women in clinical trials and the lack of sex-specific outcome data resulted in a demand from funding bodies, governmental agencies, scientific journal editors, and the general public for sex to be considered in research (National Institutes of Health, 2001; Heidari *et al.*, 2016; Clayton, 2018; Lee, 2018; Canadian Institutes of Health Research, 2019; Khramtsova, Davis and Stranger, 2019; 1.6 miljonerklubben, 2020). However, sex-specific effects are still commonly not studied, reported, or released, and they remain poorly understood (Beery and Zucker, 2011; Geller *et al.*, 2018; Lee, 2018; CARDIoGRAMplusC4D Consortium, 2020; DIAGRAM, 2020). Yet, it is crucial that we better understand any sex-specific causes and consequences of diseases if we are to determine the best preventative, prognostic, and therapeutic strategies for each sex.

1.2. The obesity pandemic – a problem for both men and women

1.2.1. The rise in obesity rates

It has been estimated that 12% of the global adult population – 600 million people – were obese (body mass index (BMI)≥30 kg/m^2) in 2015 (GBD 2015 Obesity Collaborators *et al.*, 2017). The prevalence of obesity has been increasing steadily over the past decades in most countries, and the prevalence is now exceeding one third of adults in several countries (for example the US and

Egypt) (NCD Risk Factor Collaboration (NCD-RisC), 2016b; GBD 2015 Obesity Collaborators *et al.*, 2017). Yet, obesity prevalence does not only differ between countries, but also varies across different age- and socioeconomic strata, as well as between the sexes (NCD Risk Factor Collaboration (NCD-RisC), 2016b; GBD 2015 Obesity Collaborators *et al.*, 2017). In general, obesity rates are higher in women than in men, but sex differences in the prevalence of obesity are dependent on age and country (NCD Risk Factor Collaboration (NCD-RisC), 2016b; GBD 2015 Obesity Collaborators *et al.*, 2017; The World Health Organization, 2020c). For example, obesity rates are higher in boys than girls globally, and men are more likely than women to be obese in certain countries (for example Sweden) (NCD Risk Factor Collaboration (NCD-RisC), 2016b; GBD 2015 Obesity Collaborators *et al.*, 2017; The World Health Organization, 2020c).

1.2.2. Measures of obesity

The World Health Organization defines overweight and obesity as *"abnormal or excessive fat accumulation that may impair health"* (The World Health Organization, 2020b), but how can fat and fat distribution be measured?

Total body fat can be estimated using for example dual-energy X-ray absorptiometry and bioimpedance analysis, but most larger epidemiological studies have used BMI as a proxy for total fat mass (The World Health Organization, 2000b; Völgyi *et al.*, 2008; Prospective Studies Collaboration, 2009). BMI has been criticised for having low sensitivity for identifying specific individuals at risk of obesity-associated diseases (Romero-Corral *et al.*, 2008; Okorodudu *et al.*, 2010). Yet, depending on variables such as age and sex, BMI correlates with total body fat at approximately r=0.7-0.9, with a tendency towards a higher correlation in women (Seidell *et al.*, 1987; Ranasinghe *et al.*, 2013; Akindele, Phillips and Igumbor, 2016; Bradbury *et al.*, 2017). The high correlation between BMI and body fat, in combination with its ease of measurement, has resulted in BMI being commonly measured in the clinic and in large studies of obesity (Prospective Studies Collaboration, 2009; Locke *et al.*, 2015).

Whereas BMI is a good proxy for excess fat accumulation, it does not capture differences in body fat distribution well (Ranasinghe *et al.*, 2013; Akindele, Phillips and Igumbor, 2016; Bradbury *et al.*, 2017). Intra-abdominal fat has been increasingly recognised as an independent risk factor for disease (The World Health Organization, 2000b; Pischon *et al.*, 2008). Intra-abdominal fat and body fat distribution can be measured using for example magnetic resonance imaging or computer tomography, but at a high cost as well as a potential risk of harm due to radiation in the case of computer tomography (The World Health Organization, 2000b). Instead, easily-measured anthropometric measurements, such as waist circumference and waist-hip ratio (WHR), are commonly preferred in epidemiological studies (Dagenais *et al.*, 2005; Dhana *et al.*, 2016). WHR have been shown to have a high correlation with intra-abdominal fat (*r* approximately 0.55-0.75), with potentially a higher positive correlation in men than in women (Ashwell, Cole and Dixon, 1985; Seidell *et al.*, 1987). Men tend to have a higher amount of intra-abdominal fat than women, as well as a higher WHR, and the World Health Organization recommends higher WHR cut-off values for identifying individuals at risk of metabolic complications for men (≥0.90) than for women (≥0.85) (Schreiner *et al.*, 1996; Demerath *et al.*, 2007; The World Health Organization, 2008).

1.2.3. A short background to pathophysiological mechanisms of obesity

According to the United States Department of Agriculture, calories consumed per capita have increased by almost 500 calories per day between 1970 to 2010 in the US (United States Department of Agriculture Economic Research Service, 2020). It is well established that *"the fundamental cause of obesity and overweight is an energy imbalance between calories consumed and calories expended"*, as described by the World Health Organization (The World Health Organization, 2020b). Studies have indicated that increased food availability and consumption could very well explain the rise in obesity (Hall *et al.*, 2009; Swinburn, Sacks and Ravussin, 2009).

Yet, individual susceptibility to obesity is also highly dependent on hereditary biological factors (Pedersen, Lichtenstein and Svedberg, 2002; Young *et al.*, 2018).

Physiologically, the brain and the hypothalamus are recognised as key regulators of energy homeostasis (Anand and Brobeck, 1951; Locke *et al.*, 2015; Yengo *et al.*, 2018). Genetic evidence and animal experiments have highlighted the leptin-melanocortin pathway as one of the most important regulators of appetite and energy balance (Zhang *et al.*, 1994; Farooqi *et al.*, 2000, 2003; Locke *et al.*, 2015; Yengo *et al.*, 2018). In brief, the hormone leptin is secreted by fat tissue to reflect energy balance and binds to leptin receptors on cells in the brain and the hypothalamus, which activates a complex downstream cascade that regulates appetite and energy homeostasis (Zhang *et al.*, 1994; Ghamari-Langroudi, Srisai and Cone, 2011; Baldini and Phelan, 2019). However, whereas the hypothalamus and the brain seem to be the main regulators of overall obesity and BMI, body fat distribution seems to be more related to adipogenesis and adipocyte regulation (Anand and Brobeck, 1951; Locke *et al.*, 2015; Shungin *et al.*, 2015; Yengo *et al.*, 2018). It is increasingly recognised that subcutaneous and visceral fat are fundamentally different in their endocrinological and metabolic effects, with primarily visceral fat being associated with detrimental effects on insulin resistance, dyslipidaemia, inflammation, and other metabolic risk factors (Fujioka *et al.*, 1987; Despres *et al.*, 1989; Fried, Bunkin and Greenberg, 1998; Wajchenberg, 2000; Fox *et al.*, 2007; Cartier *et al.*, 2009; Ibrahim, 2010; Tchernof and Després, 2013; Hwang *et al.*, 2015).

1.2.4. The global burden of obesity

1.2.4.1. Obesity and leading causes of death

Observational studies have linked excess weight to several of the leading cardiovascular causes of death, including coronary artery disease (CAD), stroke, and diabetes (Prospective Studies Collaboration, 2009; GBD 2015 Obesity Collaborators *et al.*, 2017; The World Health Organization, 2018b, 2020b). In addition, obesity is also highlighted as a risk factor for common

cancers, such as colorectal and breast cancer (The World Health Organization, 2020b). Yet, the extent to which the effects of obesity are mediated by overall fat versus fat distribution, interact with age, differ between men and women, or are only relevant for certain subtypes of disease is oftentimes less clear. For example, even though BMI and WHR correlate (*r* approximately 0.38-0.60) (Taylor *et al.*, 2010), several observational studies have demonstrated that BMI and WHR can be independently linked to higher risk of cardiometabolic disease (Dagenais *et al.*, 2005; Meisinger *et al.*, 2006; Pischon *et al.*, 2008; Rost *et al.*, 2018). There may also be sex differences in the effects of BMI and WHR on cardiovascular disease (Dagenais *et al.*, 2005; Li *et al.*, 2006; Abete *et al.*, 2015; Dhana *et al.*, 2016; Rost *et al.*, 2018). For example, BMI and WHR have both been implicated in stroke pathophysiology (Hu *et al.*, 2007). Yet, whereas one study found that BMI was associated with increased risk of ischaemic stroke in both sexes, BMI was only associated with haemorrhagic stroke in women, and WHR was only associated with ischaemic stroke in men (Hu *et al.*, 2007).

Although it is difficult to firmly establish causal relationships between obesity traits and cardiometabolic disease on a more granular level, the effect of obesity on global health is pervasive. Official estimates from the World Health Organization suggest that 2.8 million people die due to being overweight (BMI≥25) or obese each year (The World Health Organization, 2020b). Another study estimates that excess (higher than ideal) BMI contributes to four million deaths each year, with as many as 40% of the excess deaths occurring in people that are below the obesity threshold (GBD 2015 Obesity Collaborators *et al.*, 2017). In spite of medical advances in the treatment of diseases, global mortality rates due to high BMI increased by an estimated 28% between 1990-2015 (GBD 2015 Obesity Collaborators *et al.*, 2017). Being overweight or obese also has detrimental effects on quality of life – it has been estimated that up to 5% of disability-adjusted life years in adults can be attributed to higher-than-optimal BMI (defined as ≥22.5 kg/m^2 in the study) (GBD 2015 Obesity Collaborators *et al.*, 2017).

1.2.4.2. Obesity and female reproductive diseases

1.2.4.2.1. Obesity may increase the risk of polycystic ovary syndrome and miscarriage

Obesity has also been linked to several common female reproductive diseases, such as miscarriage (Boots and Stephenson, 2011; Cavalcante *et al.*, 2019) and polycystic ovary syndrome (PCOS) (Yildiz, Knochenhauer and Azziz, 2008; Fauser *et al.*, 2012; Brower *et al.*, 2019). However, the relationship between obesity and these female reproductive diseases is complicated and poorly understood. For example, while studies have indicated that obesity is a causal risk factor for PCOS and weight loss is a fundamental part of the treatment of the disease, PCOS may also increase the risk of obesity (Yildiz, Knochenhauer and Azziz, 2008; Moran *et al.*, 2009; Fauser *et al.*, 2012; Brower *et al.*, 2019; Lim *et al.*, 2019). There have therefore been concerns that observational studies investigating a causal relationship between obesity and PCOS may have been confounded by reverse causation or environmental confounders (Yildiz, Knochenhauer and Azziz, 2008; Fauser *et al.*, 2012; Brower *et al.*, 2019).

1.2.4.2.2 Polycystic ovary syndrome – a multi-facetted disease with severe consequences

Up to 10% of women of reproductive age may be afflicted by PCOS (Bozdag *et al.*, 2016), an endocrine, metabolic and reproductive disease in women (Fauser *et al.*, 2012). PCOS pathophysiology is still largely unclear, although one mechanism may be disrupted gonadotropin signalling that disturbs normal follicular development and ovulation (Teede, Deeks and Moran, 2010; Azziz *et al.*, 2016). Women are commonly diagnosed using the Rotterdam criteria, where diagnosis is based on a woman fulfilling two out of three criteria:

a) Oligo- or anovulation
b) Clinical and/or biochemical signs of hyperandrogenism
c) Polycystic ovaries

as well as exclusion of other diagnoses with a similar clinical presentation (Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group, 2004). PCOS is associated with wide-ranging consequences on patient health and well-being. In addition to endocrine and

reproductive consequences (Fauser *et al.*, 2012), the disease has also been associated with T2D (Moran *et al.*, 2010), cardiometabolic risk factors (Meyer, McGrath and Teede, 2005), infertility (Joham *et al.*, 2015), depression (Damone *et al.*, 2019; Greenwood *et al.*, 2019) and a generally lower self-reported health-related quality of life (Jones *et al.*, 2008).

1.2.4.2.3. Miscarriage – a common experience for women

There are several definitions of miscarriage, but in essence it is the non-intentional loss of a foetus prior to viability in a pregnancy (The World Health Organization, 2000a; The American College of Obstetricians and Gynecologists, 2018). A study in 53,479 women found that 43% of parous women had experienced at least one miscarriage, and 6% of parous women had experienced three or more miscarriages (Cohain, Buxbaum and Mankuta, 2017). At the time of the event, a miscarriage is immediately associated with bleeding, infection (Trinder *et al.*, 2006), and death (Cantwell *et al.*, 2011). Afterwards, women are at a higher risk of depression, anxiety and suicide (Cumming *et al.*, 2007; Weng *et al.*, 2018), as well as a long-term higher risk of cardiovascular disease (Maino *et al.*, 2016).

1.2.4.2.4. The need for research into obesity-associated female reproductive diseases

Obesity-associated female reproductive diseases, such as PCOS and miscarriage, thus come with far-reaching consequences that have major impact on quality of life for women (Jones *et al.*, 2008; Boots and Stephenson, 2011; Brower *et al.*, 2019; Cavalcante *et al.*, 2019). These and other diseases that occur in women only are estimated to contribute to 5% of the total disability-adjusted life years in women, whereas only 0.7% of disability-adjusted life years in men are due to diseases occurring in men only (Buvinić *et al.*, 2011). Yet, research into PCOS and miscarriage have been hampered by small sample sizes and/or poor study quality (Aleman *et al.*, 2005; Broder-Fingert *et al.*, 2009; Sivayoganathan *et al.*, 2011; Chen, Fu and Huang, 2016; Qiao *et al.*, 2016; Pereza *et al.*, 2017; Wahabi *et al.*, 2018; Lim *et al.*, 2019). From a clinical perspective, PCOS and its consequences tend to be under-diagnosed, under-treated, and over a third of women have to wait over two years for a diagnosis (Broder-Fingert *et al.*, 2009; Sivayoganathan *et al.*, 2011; Gibson-

Helm *et al.*, 2017; Wolf *et al.*, 2018). It is thus crucial to improve the understanding of the role of obesity in these diseases and to identify new potential preventative and therapeutic measures.

1.2.5. The current state of drugs and biomarkers in obesity-related disorders

1.2.5.1. Drugs for obesity

It is evident that we need to curb the rise in obesity and mitigate its consequences. Yet, effective preventative measures and treatments are limited. Lifestyle changes can result in weight loss, but the absolute weight reduction is usually modest and most of the lost weight is regained long-term (Wadden and Frey, 1997; Anderson *et al.*, 1999, 2001; Dombrowski *et al.*, 2014). Surgical interventions likely have better long-term weight loss results than non-invasive treatments, but are accompanied with operative risks and may result in other long-term sequelae, such as malabsorption (Gloy *et al.*, 2013; Heymsfield and Wadden, 2017). In spite of a major need for efficient and safe anti-obesity drugs, drugs developed to date generally have poor efficacy and several have been troubled by safety concerns (Christensen *et al.*, 2007; James *et al.*, 2010; Dombrowski *et al.*, 2014; Khera *et al.*, 2016; Shettar, Patel and Kidambi, 2017). A meta-analysis by Khera *et al.* (2016) compared the obesity drugs that were currently approved by the US Food and Drug Administration with placebo; lorcaserin, naltrexone-bupropion, phentermine-topiramate, liraglutide, and orlistat. All drugs were associated with a larger weight loss compared to placebo, ranging from -2.6 kg for orlistat to -8.8 kg for phentermine-topiramate (Khera *et al.*, 2016). Yet, since the study was published in 2016, one of the drugs (lorcaserin, a selective serotonin 2c receptor agonist) has been withdrawn from the US market due to a potentially increased risk of cancer (Chan *et al.*, 2013; U.S. Food and Drug Administration, 2020). Another drug, naltrexone-bupropion (an opioid antagonist and an aminoketone antidepressant, respectively) is prescribed under major warnings of suicidal thoughts in the US (Takeda Pharmaceuticals America, 2014; National Institute of Diabetes and Digestive and Kidney Diseases, 2016) and is not currently recommended under the National Institute for Health and Care Excellence guidelines in the UK

because of uncertain long-term effects (National Institute for Health and Care Excellence, 2017). A third drug, phentermine-topiramate (for which the exact mechanism of action is still not completely understood) was not even approved by the European Medicines Agency due to concerns about cardiovascular, psychiatric, cognitive, and teratogenic effects (European Medicines Agency, 2013). Compared to the other drugs in the meta-analysis by Khera *et al.* (2016), the glucagon-like peptide-1 analogue liraglutide was the drug with the highest likelihood of discontinued use due to side-effects (Vilsbøll *et al.*, 2007). The fifth drug orlistat, which inhibits gastrointestinal lipases, is currently considered relatively safe (Hauptman, Jeunet and Hartmann, 1992; National Institute of Diabetes and Digestive and Kidney Diseases, 2016). However, orlistat is only associated with a modest weight loss; -2.6 kg compared to placebo in the meta-analysis by Khera *et al.* (2016). Since the study by Khera *et al.* (2016), a randomised clinical trial was published with promising results for another glucagon-like peptide-1 analogue, semaglutide, in treating obesity (Wilding *et al.*, 2021). Participants who were given semaglutide lost an average of −12.4 percentage points in body weight compared to placebo. Yet, whereas promising, semaglutide is administered subcutaneously, which may diminish its usability in the clinic, and more studies are needed to exclude the risk of long-term serious adverse events (Ingelfinger and Rosen, 2021; Wilding *et al.*, 2021). In summary, there is a need for new efficient and safe drugs to treat obesity.

1.2.5.2. Drugs for obesity-related cardiometabolic diseases

Drugs treating the *consequences* of obesity have been more successful. Patients and clinicians now have a range of treatment options for cardiometabolic diseases such as T2D and CAD. Still, many of the available T2D and CAD therapies have side-effects that reduce patient compliance and/or require careful safety monitoring (for example reviewed by Thrasher (2017) for T2D and Braun *et al.* (2018) for CAD). In addition, monotherapies of cardiometabolic diseases and their comorbidities are frequently insufficient to reach patient treatment targets, and patients often require two or even three drugs for adequate disease control (American Diabetes Association, 2010; Phung *et al.*, 2014; Thrasher, 2017; Braun, Stevens and Barstow, 2018; Noubiap *et al.*, 2019). Given the large impact of cardiometabolic diseases on patient mortality and morbidity,

these observations underscore the need for new drugs for obesity and obesity-related diseases that are safe, effective, and promote adherence (The World Health Organization, 2020b). Yet, it is also important to consider potential sex differences in drug efficacy and safety, which can be exemplified with aspirin. In 1988, preliminary data from the Physicians' Health Study suggested that aspirin could decrease the risk of myocardial infarction by 47% (The Steering Committee of the Physicians' Health Study Research, 1988; Marshall, 2005). However, the study only included men, which prompted the National Institutes of Health to launch a similar study in women, in which no protective effect of aspirin on myocardial infarction was found (Marshall, 2005; Ridker *et al.*, 2005). A later study confirmed that the net benefit of aspirin as primary prevention against cardiovascular disease is much lower in women (2.5%) compared to men (12.1%) (Selak *et al.*, 2019).

1.2.5.3. Drugs for obesity-related female reproductive diseases

Currently used drugs for PCOS are mostly targeting individual symptoms of the disease. For example, a PCOS Consensus Workshop Group (Tarlatzis *et al.*, 2008; Fauser *et al.*, 2012) recommends metformin for impaired glucose tolerance, oral contraceptive pills or antiandrogens for hirsutism, and a variety of medications for infertility such as the anti-oestrogen clomiphene citrate and the gonadotropin follicle stimulating hormone (FSH).

Treatment options to prevent miscarriage are limited, and studies have been hampered by small sample sizes and/or poor study quality (Aleman *et al.*, 2005; Rumbold *et al.*, 2011; Jurkovic, Overton and Bender-Atik, 2013; Wahabi *et al.*, 2018). A Cochrane meta-analysis (Wahabi *et al.*, 2018) found moderate evidence that progestogen may decrease the risk of miscarriage, but the authors emphasised that the included studies had small sample sizes and/or low study quality (Wahabi *et al.*, 2018). Hence, the current treatment recommendations in England are focused on ensuring a safe completion of a miscarriage, rather than stopping it (National Institute for Health and Care Excellence, 2019).

1.2.5.4. Biomarkers for obesity-related cardiometabolic diseases

Another approach to mitigate the consequences of the obesity pandemic is to identify novel biomarkers that can better predict obesity and cardiometabolic diseases, and thereby offer an opportunity for targeted intervention. The National Institute of Health and the US Food and Drug Administration defines a biomarker as *"a defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or biological responses to an exposure or intervention, including therapeutic interventions"* (FDA-NIH Biomarker Working Group, 2016).

Biomarkers are already commonly used to predict and diagnose cardiometabolic diseases, as well as to evaluate treatment efficacy. For example, HbA1c can be used both to identify individuals at high risk of developing diabetes, to diagnose diabetes, and to evaluate treatment success (American Diabetes Association, 2010). Lifestyle changes in individuals at a high risk of T2D can decrease the risk of progression to overt T2D (Galaviz *et al.*, 2018), which has resulted in a large-scale national program to detect and treat individuals at high risk of T2D in England (Valabhji *et al.*, 2020). Yet, there can be sex differences in biomarker levels and utility as well. For example, troponin is commonly used as a biomarker for CAD, and values over the 99th percentile are indicative of myocardial ischaemia (Mendis *et al.*, 2011; Sobhani *et al.*, 2018). Yet, in spite of women having lower troponin levels, it is not until recently that sex-specific cut-off values have started to be recommended to prevent under-diagnosis of CAD in women (Säfström, Lindahl and Swahn, 2000; Mendis *et al.*, 2011; Kimenai *et al.*, 2018; Sobhani *et al.*, 2018; Apple *et al.*, 2020; National Institute for Health and Care Excellence, 2020).

1.2.5.5. Biomarkers for obesity-related female reproductive diseases

PCOS can be difficult to diagnose; one survey study found that 34% of women had to wait more than two years for diagnosis and that 47% had to see three or more health professionals before diagnosis (Gibson-Helm *et al.*, 2017). There has therefore been a relatively large interest in trying to identify biomarkers to aid screening and diagnosis of PCOS, as outlined below.

Hyperandrogenism – clinical or biochemical – is one of the diagnosis criteria for PCOS, but biochemical hyperandrogenism can be difficult to define (Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group, 2004). There is a lack of reference ranges for androgens where age and obesity level are considered (Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group, 2004). Kits to measure serum testosterone values have a large between-kit variability (Boots *et al.*, 1998), and many women with PCOS have normal androgen levels (Balen *et al.*, 1995; Knochenhauer *et al.*, 1998; Asunción *et al.*, 2000). In spite of these limitations, a PCOS consensus workshop group noted that free testosterone or free testosterone/androgen index could be used to assess biochemical hyperandrogenism in PCOS (Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group, 2004).

In addition to measuring androgen levels, there has also been a long-lasting debate about the potential use of female sex hormones to diagnose PCOS, in particular luteinizing hormone (LH), an LH/FSH ratio, and anti-Müellerian hormone (Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group, 2004; Cho *et al.*, 2006; Barth *et al.*, 2010; Tola *et al.*, 2018; Le *et al.*, 2019). Women with PCOS often have higher LH levels and lower FSH levels than controls, but studies have been contradictory regarding its usefulness in diagnosing PCOS (Cho *et al.*, 2006; Barth *et al.*, 2010; Tola *et al.*, 2018; Le *et al.*, 2019). The difficulty in using female sex hormones to diagnose PCOS may, in part, be due their changed levels across menstrual and circadian cycles (Yeung *et al.*, 2013; Rahman *et al.*, 2019). Further complicating matters is that obesity is also associated with changed LH and FSH levels across the menstrual cycle (Yeung *et al.*, 2013).

Female sex hormones have also been considered as biomarkers in pregnancy and miscarriage, of which the most well-known one is likely human chorionic gonadotropin. Human chorionic gonadotropin increases rapidly in early pregnancy (Pittaway, Reish and Wentz, 1985) and is measured in commercial pregnancy tests (Clearblue, 2020). A slower increase in human chorionic gonadotropin levels is associated with miscarriage, but inaccurate dating of pregnancies limits its use in predicting miscarriage (Lower and Yovich, 1992). Still, declining levels of human chorionic

gonadotropin are used to differentiate between miscarriages and pregnancies of unknown location that may require intervention (Condous *et al.*, 2006). Another biomarker relevant for miscarriage is progesterone. Low serum progesterone is strongly linked to a non-viable pregnancy (Verhaegen *et al.*, 2012). However, as low levels of progesterone can also be present in viable pregnancies, it cannot be used as a definitive test in isolation (Cordina *et al.*, 2011; Jurkovic, Overton and Bender-Atik, 2013).

1.3. Obesity and genetics

1.3.1. An introduction to genome-wide association studies

1.3.1.1. Basic concepts and challenges of genome-wide association studies

In 2001, the International Human Genome Project Consortium and Celera Genomics published the first drafts of the human genome sequence (Lander *et al.*, 2001; Venter *et al.*, 2001). The International HapMap Project was launched the following year, which greatly expanded our knowledge of how genetic variants correlate with each other across the genome and resulted in a public database of common genetic variations across the genome (Belmont *et al.*, 2005). These landmark events, together with later technological and methodological advances, enabled GWASs and the emergence of the field of genetic epidemiology (Butler, 2001).

A GWAS quantifies the associations between genetic variants, usually single nucleotide polymorphisms (SNPs), and a trait or disease of interest, usually under the assumption of an additive allelic effect and providing effect estimates per additional allele (Marees *et al.*, 2018). Yet, a genotype-trait estimate from a classic GWAS will not only include the direct effect of the allele, but also any indirect and confounding effects from, for example, population stratification, assortative mating, and gene-environment interactions (Sonestedt *et al.*, 2009; Kong *et al.*, 2018; Young *et al.*, 2019; Brumpton *et al.*, 2020; Morris *et al.*, 2020). Of these, population stratification in particular has received a lot of attention, and it has likely been a source of error in large GWASs

(Berg *et al.*, 2019). Population stratification occurs when a genotype frequency and the phenotype under investigation systematically vary in subgroups of the study population, which will cause the genotype to seem associated with the phenotype (Marees *et al.*, 2018). For example, Knowler *et al.* (1988) found that the immunoglobulin G haplotype $Gm^{3;5,13,14}$ was seemingly associated with T2D risk in participants of Pima and Papago ancestry, but demonstrated that the effect was driven by both the haplotype frequency and the T2D prevalence being associated with degree of European ancestry. Another example was the apparent association between a genetic variant in the lactase gene and height, which was no longer present after geographical stratification of the sample (Campbell *et al.*, 2005).

Another important aspect of GWASs is linkage disequilibrium (LD), the non-random associations between alleles at different genomic positions (Shifman *et al.*, 2003). In general, close genetic variants are in higher LD, and close SNPs can be perfectly correlated with each other in a sample (Shifman *et al.*, 2003). Alleles at genomic positions that have not been genotyped in a dataset can thereby be imputed – essentially inferred through their correlation with genotyped SNPs – by using more densely genotyped/sequenced reference samples of a similar ancestral background (Belmont *et al.*, 2005; Astle and Balding, 2009; Howie, Donnelly and Marchini, 2009). In addition, even if the actual causal SNP have not been genotyped or imputed, tagging variants may lead to the associated locus still being identified (Astle and Balding, 2009). Yet, LD can also make it difficult to determine the actual causal SNP in fine-mapping efforts, as many SNPs in high LD may show very similar strengths of association with the trait under investigation (Mahajan, Taliun, *et al.*, 2018).

Another potential issue in GWASs is collider bias. A collider is a variable that is affected by both the independent and dependent variable in a regression, and adjustment for a collider can bias effect estimates (Aschard *et al.*, 2015; Day *et al.*, 2016). Usually, adjustments are done to account for confounders, as well as to increase power by reducing residual variance, but in GWASs adjustments for heritable confounders can introduce collider bias (Aschard *et al.*, 2015). Day *et al.*

(2016) illustrate this by performing a GWAS of sex and adjusting for height. In the study, height becomes a collider as it is affected by both sex and height-associated genetic variants (Day *et al.*, 2016). The result is an apparent association between the height-associated genetic variants and sex (Day *et al.*, 2016). In a similar manner, participation (or selection) bias dependent on phenotypes can result in spurious associations between traits, which for example can be seen in The Avon Longitudinal Study of Parents and Children (Munafò *et al.*, 2018). In the study, mothers were less likely to remain in the study if they had a higher BMI or smoked more (Munafò *et al.*, 2018). If both high BMI and smoking affect continued participation in the study, participation effectively becomes a collider and could result in a biased estimate of the association between BMI and smoking (Munafò *et al.*, 2018).

1.3.1.2. The genetic architecture of common complex traits

In 2009, The International Schizophrenia Consortium demonstrated that polygenic risk scores of schizophrenia – computed as an effect estimate-weighted sum of the schizophrenia-increasing alleles of each individual – were associated with schizophrenia, even when the included SNPs did not surpass the GWAS-significance threshold (International Schizophrenia Consortium *et al.*, 2009). As GWASs of common, complex traits gained larger sample sizes, more genetic loci were identified, which contributed to the idea of an allelic spectrum (or allelic series) of common disease (Locke *et al.*, 2015; Shungin *et al.*, 2015; Yengo *et al.*, 2018; Barroso and McCarthy, 2019; Pulit *et al.*, 2019). In brief, the idea is that genetic trait variance for common, complex disease is primarily made up of rare high-impact alleles, low-frequency medium-impact alleles, and low-impact common alleles (Barroso and McCarthy, 2019). For example, rare variants in the leptin-melanocortin pathway have individually been linked to severe obesity (Farooqi *et al.*, 2000, 2003) whereas common variants in the pathway have more modest effects on obesity (Yengo *et al.*, 2018; Pulit *et al.*, 2019). However, the genetic architecture can vary dramatically between traits and diseases. For example, contrary to common, non-syndromic obesity which have a polygenic background (Farooqi *et al.*, 2000, 2003; Yengo *et al.*, 2018; Pulit *et al.*, 2019), cystic fibrosis is a

wholly monogenic disease caused by mutations in a single gene (Riordan *et al.*, 1989; Rommens *et al.*, 1989; Bobadilla *et al.*, 2002).

1.3.1.3. Sex in genome-wide association studies

Women outnumber men in eight out of ten of the most frequently used datasets for GWASs (Mills and Rahal, 2019). For some studies, this is due to study design, as exemplified in the Nurses' Health Study (Belanger *et al.*, 1978). However, there can also be sex differences in how likely an individual is to participate in a study. For example, the participation rate was higher in invited women (6.4%) than men (5.1%) in the UK Biobank (UKBB), a large, population-based cohort study commonly used in recent GWASs (Sudlow *et al.*, 2015; Fry *et al.*, 2017; Mahajan, Taliun, *et al.*, 2018; Pulit *et al.*, 2019). Yet, similar to randomised controlled trials, many of the large GWAS consortia have only presented sex-combined outcome data and/or only released sex-combined summary statistics (Nelson *et al.*, 2017; GEFOS Consortium, 2018; Mahajan, Taliun, *et al.*, 2018; Morris *et al.*, 2019; CARDIoGRAMplusC4D Consortium, 2020; DIAGRAM, 2020). Historically, this likely reflects a lack of power to identify associations in smaller sex-stratified sub-cohorts. Still, even relatively recent GWASs with sample sizes in the hundreds of thousands – and hundreds of discovered loci – have only released sex-combined data (Nelson *et al.*, 2017; GEFOS Consortium, 2018; Mahajan, Taliun, *et al.*, 2018; Morris *et al.*, 2019; CARDIoGRAMplusC4D Consortium, 2020; DIAGRAM, 2020).

1.3.2. The role of genetics in obesity

1.3.2.1. Genome-wide association studies of obesity

In spite of the challenges in performing GWASs, approximately a thousand independent genetic variants associated with obesity have been identified, primarily thanks to efforts by the Genetic Investigation of ANthropometric Traits (GIANT) consortium (Locke *et al.*, 2015; Shungin *et al.*, 2015; Yengo *et al.*, 2018; Pulit *et al.*, 2019). In 2015, the GIANT consortium performed GWASs of anthropometric traits including up to 339,224 participants and identified or replicated over a hundred genetic loci associated with obesity traits (Locke *et al.*, 2015; Shungin *et al.*, 2015). Both

sex-specific and sex-combined summary statistics for BMI, WHR, as well as other anthropometric measures with and without adjustment for BMI were made publicly available, enabling other researchers to build on these findings (Locke *et al.*, 2015; Shungin *et al.*, 2015). In 2018, the GIANT consortium released new summary statistics that included data from the UKBB (Sudlow *et al.*, 2015; Yengo *et al.*, 2018; Pulit *et al.*, 2019). For BMI, the new, updated GWASs included over 700,000 samples and almost a thousand near-independent genetic variants were found to be associated with BMI (Yengo *et al.*, 2018). For WHR adjusted for BMI (WHRadjBMI), the new GWAS included almost 700,000 samples and with 346 genetic loci surpassing the multiple testing threshold of $P<5\times10^{-9}$ (Pulit *et al.*, 2019). Together, the most associated SNPs for each trait explained approximately 6.0% of the trait variance for BMI and 3.9% for WHRadjBMI, respectively (Yengo *et al.*, 2018; Pulit *et al.*, 2019). Yet, a trait variance explained of 6.0% and 3.9% is far from the heritability estimates for these traits provided by twin-studies. Traditional twin studies have estimated the heritability of BMI to approximately 65%, but conservative estimates using methods less prone to environmental bias suggest the heritability may be as low as 29% (Pedersen, Lichtenstein and Svedberg, 2002; Young *et al.*, 2018).

1.3.2.2. Sex differences in the genetics of obesity

Sex differences in genetic architecture seem to be trait-dependent. The study by Pulit *et al.* (2019) found evidence for the SNP-based heritability (h_g^2), computed using the restricted maximum likelihood method (Loh *et al.*, 2015), to be higher in women than in men for both WHR (women h_g^2=27.7%; men h_g^2=21.3%; P-value for a sex difference ($P_{sexdiff}$)=1.2×10^{-26}) and WHRadjBMI (women h_g^2=29.3%; men h_g^2=16.6%; $P_{sexdiff}$=4.0×10^{-103}). However, there was no evidence of a sex difference in the SNP-based heritability estimates for BMI (women h_g^2=33.5%; men h_g^2=34.7%; $P_{sexdiff}$=0.07) (Pulit *et al.*, 2019). Another study by Ge *et al.* (2017) found that 13 out of 14 traits with sex-differential SNP-based heritability (out of 551 complex traits investigated) had higher heritability estimates in women. Of these 13 traits with a higher sex-differential SNP-based heritability in women, half were anthropometric and/or blood pressure traits, suggesting that there

may be inherent sex differences in the genetic architecture of obesity and cardiometabolic traits (Ge *et al.*, 2017).

Furthermore, Pulit *et al.* (2019) identified 266 genetic loci associated ($P<5\times10^{-9}$) with WHRadjBMI in women, but only 91 loci associated with WHRadjBMI in men. While a part of this sex difference could be explained by the larger sample size in women (379,501 women versus 315,284 men), 105 index SNPs were also found to have sex-differential effect estimates for WHRadjBMI (Pulit *et al.*, 2019). Over 90% of these 105 SNPs had higher effect estimates in women than in men (Pulit *et al.*, 2019). On the contrary, there were almost no sex-differential index SNPs for BMI; only four of the index SNPs identified in Europeans had evidence of sex-differential effect estimates (Pulit *et al.*, 2019). Thus, it is clear that the genetic architecture, as well as individual SNP effect estimates, may differ between men and women in obesity.

1.4. Using genetic epidemiology to tease out relationships between traits

1.4.1. Leveraging genetic associations to explore associations between traits

It is clear that genetic variation affects individual susceptibility for many common, complex traits and diseases (International Schizophrenia Consortium *et al.*, 2009; Yengo *et al.*, 2018; Buniello *et al.*, 2019; Pulit *et al.*, 2019). Robust associations between genetic variants and a phenotype can be leveraged to clarify relationships between traits of interest, such as diseases, risk factors, and intermediate cellular phenotypes (Davey Smith and Ebrahim, 2003; Giambartolomei *et al.*, 2014; Hormozdiari *et al.*, 2016; Burgess, Davey Smith, *et al.*, 2020). A plethora of methods have been developed with the aim to clarify relationships between traits, with two of the most common types being various forms of Mendelian randomisation (MR) and colocalisation methods (Davey Smith

and Ebrahim, 2003; Giambartolomei *et al.*, 2014; Hormozdiari *et al.*, 2016; Burgess, Davey Smith, *et al.*, 2020).

1.4.2. Mendelian randomisation

1.4.2.1. A basic introduction to Mendelian randomisation

Conventional observational studies may be biased by reverse causation and confounding (Davey Smith and Ebrahim, 2003). There are several examples of how substantial observational associations between an exposure and a disease could not be replicated in randomised clinical trials. For example, a large prospective study of almost 20,000 participants found vitamin C to be protective against CAD (Khaw *et al.*, 2001). However, this finding could not be replicated in a later randomised clinical trial (Collins *et al.*, 2002). Similarly, while observational studies have indicated that higher vitamin E levels are associated with lower cardiovascular mortality, intervention studies have not confirmed a protective effect (Hooper, Ness and Smith, 2001).

MR offers another approach to evaluate causal relationships by using genetic associations. In 1865, Gregor Mendel suggested that the inheritance of two traits were independent events (Mendel, 1865). Today, we know that this is not always true, for example due to LD or pleiotropy, that is, when a genetic locus is affecting two seemingly unrelated traits (Burgess and Thompson, 2015; Sun *et al.*, 2018). Yet, the principle of random assortment of genotypes at gamete formation is fundamental for MR (Davey Smith and Ebrahim, 2003). One of the earliest suggestions of applying MR methodology to determine causal relationships was made in 1986 by Katan (1986). He was concerned with the apparent inverse association between cholesterol and cancer in observational studies, and suggested a simple method to avoid bias from reverse causation and/or confounders (McMichael *et al.*, 1984; Katan, 1986). In brief, genetic variants in the apolipoprotein E gene were known to be associated with cholesterol (Utermann *et al.*, 1984). If genetic variants that were associated with lower cholesterol were relatively more common in cancer patients than controls, Katan (1986) suggested that this would strengthen the hypothesis that lower cholesterol

levels increased the risk of cancer. However, if no relative difference in cholesterol-associated variants was seen between cancer patients and healthy controls, that would suggest that the observational finding was spurious (Katan, 1986).

Katan's idea remains the essence of MR analysis. In MR studies, randomly allocated genetic variants that predispose for an exposure are used as proxies (or instruments) for the exposure itself (Greenland, 2000; Davey Smith and Ebrahim, 2003). If such genetic instruments for an exposure are also associated with an outcome – and only affect the outcome through the exposure – a causal relationship between the exposure and the outcome can be established (Davey Smith and Ebrahim, 2003). Still, for MR analyses to be valid, several assumptions must be met, as discussed below.

1.4.2.2. Mendelian randomisation assumptions

The three core assumptions of MR state that (Greenland, 2000; Martens *et al.*, 2006; Burgess and Thompson, 2015):

a) The genetic variant(s) must be associated with the exposure.
b) The genetic variant(s) cannot be associated with any confounders in the exposure-outcome relationship.
c) The genetic variant(s) only affect the outcome through the exposure.

Of these assumptions, only the first is directly verifiable (Burgess and Thompson, 2015; Burgess, Davey Smith, *et al.*, 2020). Whereas the second and third assumptions can be falsified, they cannot be verified as that would require testing of all potential confounders in the exposure-outcome relationship (Swanson and Hernán, 2013; Burgess and Thompson, 2015; Burgess, Davey Smith, *et al.*, 2020). However, even if an association between an instrument and a potential confounder is identified (known as pleiotropy), that does not per se mean that the instrument is invalid (Fig 1.1). Pleiotropic effects only invalidate the MR assumptions if they are not mediated by the exposure, known as horizontal pleiotropy (Davey Smith and Hemani, 2014; Davies, Holmes and Davey Smith, 2018). Vertical pleiotropy – in which the effects on the confounder are through the

exposure – are generally not a problem (Davey Smith and Hemani, 2014; Davies, Holmes and Davey Smith, 2018).

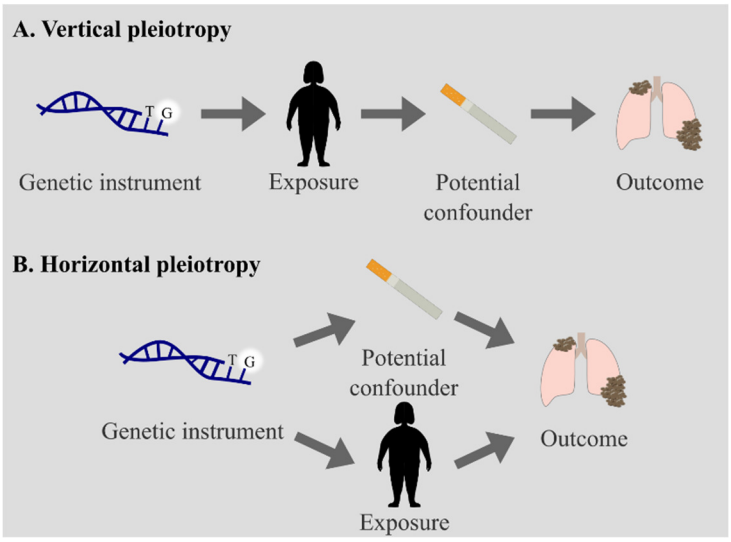


Figure 1.1. Illustration of vertical versus horizontal pleiotropy. In MR, genetic variants associated with an exposure (in this example obesity) should only affect the outcome (in this example lung cancer) through the exposure. A) In vertical pleiotropy, the association between the genetic variants and the potential confounder (in this example smoking) is only mediated through the effect on obesity, and thus the MR assumptions hold. B) In horizontal pleiotropy, the genetic variants are associated with a confounder (smoking) of the exposure-outcome relationship. As the effect is not mediated through the exposure, the MR assumptions for the genetic instrument are violated (Davey Smith and Hemani, 2014; Davies, Holmes and Davey Smith, 2018).

In addition to the three core assumptions, correct causal effect estimation also requires assumptions about homogeneity or monotonicity (Swanson and Hernán, 2013; Burgess and Thompson, 2015). In brief, the homogeneity assumption is that all instruments identify the same causal effect of the exposure on the outcome (Swanson and Hernán, 2013; Burgess and Thompson, 2015). The monotonicity assumption can be seen as a relaxation of the homogeneity assumption, i.e., that genetic instruments have the same directional (or zero) effect on the exposure in all samples (Swanson and Hernán, 2013; Burgess and Thompson, 2015).

In the absence of violations of the MR assumptions, MR findings of a causal effect should be unbiased by confounders (Davey Smith and Ebrahim, 2003). In addition, if the MR assumptions hold, reverse causation is unlikely as genotypes are allocated at conception (Davey Smith and Ebrahim, 2003). Yet, there are several reasons why an MR estimate of a causal effect may differ from real-life effects of intervening on an exposure. For example, MR estimates captures the

effects from lifelong perturbations of the exposure on the outcome (Burgess and Thompson, 2015; Davies, Holmes and Davey Smith, 2018). Effect estimates from intervention studies, performed later in life, may therefore differ from the MR estimates (Burgess and Thompson, 2015; Davies, Holmes and Davey Smith, 2018). In addition, developmental disruptions caused by genotypes can be buffered against through compensatory mechanisms, known as canalisation (Waddington, 1942; Lawlor *et al.*, 2008). Thus, it has been recommended to not interpret MR estimates too literally (Burgess and Thompson, 2015; Burgess *et al.*, 2017).

1.4.2.3. An overview of Mendelian randomisation study design

MR studies can be conducted either in one sample (where genetic instruments, the exposure, and the outcome is measured in a single sample) or two samples (where instrument-exposure associations are measured in one sample and instrument-outcome associations in another sample), or partly overlapping samples (Burgess and Thompson, 2015; Burgess *et al.*, 2015; Burgess, Davies and Thompson, 2016; Burgess, Davey Smith, *et al.*, 2020). In general, a one sample study has the advantage of ensuring that all data come from the same underlying population, but could be at risk of weak instrument bias causing false positive results (see section 1.4.2.8.1) (Burgess, Davey Smith, *et al.*, 2020). On the contrary, while two sample MR studies may benefit from larger sample sizes for estimating the instrument-exposure and instrument-outcome effect estimates, they may be at a higher risk of bias from the study samples not coming from the same underlying population (Burgess, Davey Smith, *et al.*, 2020). MR methods have been developed both for individual-level data and for summary-level data (Burgess and Thompson, 2015; Burgess, Davies and Thompson, 2016). Most MR methods assume that the included genetic instruments are uncorrelated with each other (Burgess, Dudbridge and Thompson, 2016; Burgess, Davey Smith, *et al.*, 2020). MR methods that allow for inclusion of correlated instruments exist (Burgess, Dudbridge and Thompson, 2016), however the methods described in this overview assume that the genetic instruments are uncorrelated.

1.4.2.4. The release of Mendelian randomisation guidelines

A plethora of MR methods has been developed in the last decade, all with their own strengths and limitations (Burgess, Butterworth and Thompson, 2013; Bowden *et al.*, 2016; Hartwig, Smith and Bowden, 2017; Verbanck *et al.*, 2018; Qi and Chatterjee, 2019; Rees *et al.*, 2019; Burgess, Davey Smith, *et al.*, 2020; Burgess, Foley, *et al.*, 2020). In spite of the large interest in developing new MR methods, the applied MR field has been plagued by poor reporting standards (Boef, Dekkers and Le Cessie, 2015). To promote a higher standard in MR studies, leading MR researchers published recommendations in 2019, which were further updated in 2020 (Burgess, Davey Smith, *et al.*, 2020). The authors of the MR-guidelines concede that estimates can vary considerably between methods, and recommend researchers to use several methods with different assumptions to explore causal relationships (Burgess, Davey Smith, *et al.*, 2020). If different methods give consistent effect directions and similar effect estimates, evidence of a causal effect is strengthened (Burgess, Davey Smith, *et al.*, 2020).

1.4.2.5. Basic Mendelian randomisation methods

A genetic instrument that is both associated with the exposure as well as the outcome provides evidence of a causal effect (if the MR assumptions hold) (Davey Smith and Ebrahim, 2003; Burgess *et al.*, 2015). Yet, while simple, this approach does not provide any estimates of the causal effect (Burgess *et al.*, 2015).

One of the most basic (but still widely used) MR methods is the ratio (or Wald) method (Boef, Dekkers and Le Cessie, 2015; Burgess *et al.*, 2015; Burgess, Davey Smith, *et al.*, 2020). The ratio method is performed by dividing the instrument-outcome estimate by the instrument-exposure estimate (Burgess and Thompson, 2015; Burgess, Davey Smith, *et al.*, 2020). However, the ratio method can only be applied in the single-instrument case (Burgess and Thompson, 2015; Burgess, Davey Smith, *et al.*, 2020). For the multiple-instrument case, the inverse-variance weighted (IVW) method is commonly used instead, which in essence is an IVW-meta analysis of the individual ratio estimates for each included instrument (Boef, Dekkers and Le Cessie, 2015; Burgess and

Thompson, 2015; Burgess *et al.*, 2015; Burgess, Davey Smith, *et al.*, 2020). Equivalently, the same estimate can be obtained through the ratio method by combining the associated genetic variants into a single instrument – a genetic risk score (GRS) with the sum of each individual's exposure-increasing alleles (Burgess, Dudbridge and Thompson, 2016; Burgess, Davey Smith, *et al.*, 2020). These approaches, in turn, will essentially give the same effect estimates as the 2-stage least square method using individual-level data (Burgess and Thompson, 2015; Burgess, Dudbridge and Thompson, 2016; Burgess, Davey Smith, *et al.*, 2020).

The main limitation of the ratio method, IVW method, and the 2-stage least square method is that they assume that all included genetic variants are valid instruments (Burgess and Thompson, 2015; Bowden *et al.*, 2016; Burgess, Davey Smith, *et al.*, 2020). The methods will therefore give biased effect estimates if, for example, an instrument is affecting the outcome through a pathway that is not mediated by the exposure (Bowden, Davey Smith and Burgess, 2015; Burgess and Thompson, 2015).

1.4.2.6. Commonly used Mendelian randomisation sensitivity analyses

Other MR methods have been developed to provide unbiased results even when instruments are violating the MR assumptions (Bowden, Davey Smith and Burgess, 2015; Bowden *et al.*, 2016). One of these methods is the median approach, in which ratio estimates are computed for each instrument and the median estimate is then assumed to represent the true effect (Bowden *et al.*, 2016). Median estimates can either be computed as the simple median estimate (with all individual ratio estimates given an equal weight) or the weighted median estimate (with all individual ratio estimates given weights based on inverse variances of the ratio estimates after standardisation) (Bowden *et al.*, 2016). A major strength of the median-based methods is that they provide valid effect estimates even if up to 50% of the information comes from invalid instruments (by number of instruments for the simple median approach and by weight for the weighted median approach) (Bowden *et al.*, 2016). Similarly, a mode-based method has been proposed, which gives valid

effect estimates as long as the largest number of similar estimates come from valid instruments (Hartwig, Smith and Bowden, 2017).

Another method that is more robust to invalid instruments is MR-Egger (Bowden, Davey Smith and Burgess, 2015). The MR-Egger method gives consistent causal effect estimates even if all the instruments are invalid, as long as another assumption is fulfilled; that the instrument strength is not associated with the pleiotropic effects of the instrument (Bowden, Davey Smith and Burgess, 2015). Contrary to the IVW method, the intercept in the MR-Egger estimates is not confined to zero (Bowden, Davey Smith and Burgess, 2015). If the MR-Egger intercept deviates from zero, that is evidence of a non-zero average effect of the instruments on the outcome that is *not* mediated through the exposure, i.e. directional pleiotropy (Bowden, Davey Smith and Burgess, 2015). While the MR-Egger relaxes the core MR assumptions, a major limitation of the MR-Egger method is reduced power and less precise estimates than the IVW method (Bowden, Davey Smith and Burgess, 2015). Still, if MR-Egger estimates have consistent effect directions with the IVW estimates, and the MR-Egger intercept is not significantly different from zero, evidence of a causal exposure-outcome relationship is strengthened (Bowden, Davey Smith and Burgess, 2015).

1.4.2.7. Additional Mendelian randomisation methods

Many other methods have been developed in recent years, such as the MR-PRESSO, MR-Lasso, MR-Robust, MR-Mix, Contamination Mixture methods, and Causal Analysis Using Summary Effect estimates (CAUSE) (Verbanck *et al.*, 2018; Qi and Chatterjee, 2019; Rees *et al.*, 2019; Burgess, Foley, *et al.*, 2020; Morrison *et al.*, 2020). While these newer methods can be more robust in certain scenarios, for example in the presence of outliers, they can also be more sensitive, for example to the presence of several invalid instruments (Verbanck *et al.*, 2018; Qi and Chatterjee, 2019; Rees *et al.*, 2019; Burgess, Davey Smith, *et al.*, 2020; Burgess, Foley, *et al.*, 2020). Still, the recently published CAUSE method has performed well in simulations, and seems to lessen the risk of bias from horizontal pleiotropy while retaining high power to identify causal effects (Morrison *et al.*, 2020).

1.4.2.8. Potential biases in Mendelian randomisation

1.4.2.8.1. Overfitting and weak instrument bias

In a finite sample, instruments that are only weakly associated with the exposure may cause weak instrument bias (Staiger and Stock, 1997; Burgess and Thompson, 2015; Burgess, Davies and Thompson, 2016; Burgess, Dudbridge and Thompson, 2016). In a one-sample MR analysis, this bias will be in the direction of any observational confounding (Burgess, Davies and Thompson, 2016; Burgess, Dudbridge and Thompson, 2016). In a two-sample MR analysis, the bias will be in the direction of the null (Burgess, Davies and Thompson, 2016). In the case of an MR-analysis with partly overlapping samples, the direction and degree of bias is dependent on the overlap (Burgess, Davies and Thompson, 2016).

Weak instrument bias arises due to overfitting in the instrument-exposure, resulting in the instruments capturing chance variation in any observational confounders in addition to any true effects (Burgess and Thompson, 2013). In the case of a one-sample MR and a confounder in the exposure-outcome relationship, overfitting will then result in a correlation between the instrument-exposure and instrument-outcome regression coefficients (Burgess and Thompson, 2011; Taylor *et al.*, 2014; Burgess, Davies and Thompson, 2016). In addition, if instruments are selected based on P-values for association with the exposure, and the P-value threshold for inclusion is low, winner's curse can cause genetic variants to be selected that are not actually associated with the exposure (Taylor *et al.*, 2014). Even if the instrument is indeed associated with the exposure, winner's curse could cause instrument-exposure effect estimates to be larger than the true population estimate (Taylor *et al.*, 2014).

Intuitively, the stronger the instrument is, the more it will reflect the true trait variation explained by the instrument as opposed to chance associations with confounders (Burgess and Thompson, 2011). The relative bias – the MR estimate bias divided by the observational bias – is theoretically

close to the reciprocal of the expected value of the F-statistic (Staiger and Stock, 1997; Burgess, Davies and Thompson, 2016). The F-statistic can be computed as

$$F = \frac{N - K - 1}{K} \frac{R^2}{1 - R^2}$$

where N is the sample size, K is the number of instruments, and R^2 is the trait variance explained by the instruments (Burgess, Davies and Thompson, 2016). It is thus evident that a larger sample size, fewer instruments, and larger trait variance explained each increases the F-statistic and decreases weak instrument bias. Commonly, an F-statistic <10 has been used to denote a weak instrument, however there is no set criteria to define a weak instrument as it is a continuous phenomenon rather than a binary (Burgess and Thompson, 2013). Simulation studies have indicated that MR studies can be considerably affected by weak instrument bias, but also that such bias is primarily a problem in the case of small sample sizes and weak instruments (Burgess and Thompson, 2011, 2013; Burgess, Butterworth and Thompson, 2013; Taylor *et al.*, 2014; Burgess, Davies and Thompson, 2016; Burgess, Dudbridge and Thompson, 2016). For example, a simulation study by Burgess and Thompson (2013) found evidence of weak instrument bias when using a sample size of 3,000, but not when increasing the sample size to 30,000 (including SNPs that were associated with the exposure at $P<0.05$). Another study, by Burgess, Butterworth and Thompson (2013), found little evidence of weak instrument bias if the included instruments were associated with the exposure at $P<1\times10^{-5}$ using a sample of 5,000 participants. Besides assuring that included instruments are strongly associated with the exposure, it is also possible to decrease the bias by reducing the number of instruments into a single GRS, which in effect will increase the F-statistic (Burgess and Thompson, 2013; Burgess, Dudbridge and Thompson, 2016). A GRS is commonly computed by summing up all the exposure-increasing alleles of the genetic instruments for each individual, and is usually weighted by the effect estimate of each allele (Burgess and Thompson, 2013). To further reduce any weak instrument bias, the GRS can also be weighted using effect estimates computed in an external cohort (an externally weighted GRS) or by assigning each genetic variant a weight of 1 (an unweighted GRS) (Burgess and Thompson, 2013; Burgess, Dudbridge and Thompson, 2016). Simulations studies have indicated that unweighted

GRSs and externally weighted GRSs provide unbiased effect estimates, although potentially with reduced power in the case of unweighted or imprecisely measured externally weighted GRSs (Burgess and Thompson, 2013; Burgess, Dudbridge and Thompson, 2016).

1.4.2.8.2. Linkage disequilibrium and population effects

LD between SNPs can be both advantageous and disadvantageous in MR studies. Since LD can result in close SNPs being strongly correlated with each other (Shifman *et al.*, 2003), SNPs in high LD with each other can be used as proxies for one another (Burgess *et al.*, 2015). In fact, it is likely that the top SNP in a GWAS is not the actual causal SNP, but only tagging it (Astle and Balding, 2009). However, LD patterns can differ between ancestral populations (Shifman *et al.*, 2003; Berisa and Pickrell, 2016). If the exposure and outcome dataset differ in regards to ancestry, the instrument may, for example, be correlated with the causal SNP in the exposure dataset but not the outcome dataset, which in effect will violate the MR assumptions (Burgess *et al.*, 2015; Burgess, Davey Smith, *et al.*, 2020). In addition, it is also important the exposure and outcome study populations also are similar in regards to other characteristics, such as age, sex, and environmental exposures, so that the instrument effect estimates are accurate for both study populations.

1.4.2.9. Sex-specific Mendelian randomisation studies

MR methodology has been widely applied to tease out causal relationships between traits, in particular for cardiometabolic diseases (Corbin *et al.*, 2016; Dale *et al.*, 2017; Emdin *et al.*, 2017; Carter *et al.*, 2019; Liu *et al.*, 2019). Yet, there are only few studies that have performed sex-stratified MRs to date or investigated sex differences in causal effect estimates computed using MR.

A few studies have performed MR studies in a single sex due to the disease affecting primarily one sex, such as breast and ovarian cancer (Guo *et al.*, 2016; Shu *et al.*, 2018; Dimou *et al.*, 2019; Harris *et al.*, 2019). However, only a small proportion of studies have performed sex-stratified MRs or investigated sex differences for exposure-outcome relationships relevant to both sexes.

Lotta *et al.* (2018) investigated sex differences in the effect of body fat distribution on CAD and T2D, but did not find any evidence of a sex difference using an interaction term for sex. Huang *et al.* (2015) quantified the association between a GRS for WHR and T2D separately in men and women, but reported that the causal effect estimates were similar in both sexes. A recently published study by Ruth *et al.* (2020) performed sex-stratified MR analysis to investigate the effect of testosterone on risk of T2D. Similar to an observational study (Karakas *et al.*, 2018), Ruth *et al.* (2020) found that higher testosterone increased the risk of T2D in women, but decreased the risk of T2D in men. Yet, the large-scale MR studies that leverage protein quantitative trait loci (pQTL) data to identify new potential protein biomarkers and/or drug targets have only been performed sex-combined to date (Sun *et al.*, 2018; Yao *et al.*, 2018; Chong *et al.*, 2019; Ritchie *et al.*, 2019; Bretherick *et al.*, 2020; Folkersen *et al.*, 2020; Gudmundsdottir *et al.*, 2020; Zheng *et al.*, 2020).

1.4.3. Colocalisation

1.4.3.1. The difficulty in finding the trait-mediating mechanisms of associated genetic loci

GWASs have identified thousands of trait-associated genetic variants (Buniello *et al.*, 2019). Yet, our knowledge of how each trait-associated genetic variant contributes to a phenotype is limited. While the mediating gene is often the gene closest to the genetic locus, effect mediation can also be considerably more complicated, which can be exemplified with the obesity-associated *FTO* locus (Mahajan, Wessel, *et al.*, 2018; Barroso and McCarthy, 2019).

Early GWASs reported that SNPs in the *FTO* locus were strongly associated with BMI (Dina *et al.*, 2007; Frayling *et al.*, 2007; Scuteri *et al.*, 2007). Homozygous carriers of the obesity-increasing allele weighed on average approximately three kilograms more than participants homozygous for the other allele (Dina *et al.*, 2007; Frayling *et al.*, 2007; Scuteri *et al.*, 2007). Functional follow-up studies using overexpression and knock-out study designs suggested that *FTO* was the obesity-mediating gene at the locus (Gerken *et al.*, 2007; Fischer *et al.*, 2009; Church *et al.*, 2010). However, later studies provided robust evidence that the obesity-associated variants in the *FTO*-

locus primarily mediated the obesity-increasing effect through regulatory effects on the genes *IRX3* and *IRX5*, which are involved in adipocyte development (Smemo *et al.*, 2014; Claussnitzer *et al.*, 2015).

1.4.3.2. How colocalisation can help identify trait-mediating genes and proteins

One approach to highlight potential trait-mediating genes and proteins at a genetic locus is through colocalisation analysis. Colocalisation methods assess the likelihood of two genetic association signals being consistent with a shared causal variant, rather than two distinct causal variants in LD with each other (Giambartolomei *et al.*, 2014; Hormozdiari *et al.*, 2016). A high likelihood of colocalisation between a disease and an intermediate cellular phenotype, such as gene expression or protein levels, increases the probability that the intermediate cellular phenotype is mediating the disease risk at the disease-associated genetic locus (Giambartolomei *et al.*, 2014). Previously, colocalisation methods have highlighted potential effector genes in for example carotid plaque formation and immune-mediated diseases (Guo *et al.*, 2015; Franceschini *et al.*, 2018). Yet, evidence of colocalisation cannot be regarded as final evidence that a gene is involved in a disease. Many genes share regulatory elements and thus several genes can colocalise with a trait at a specific locus (Bhatia and Kleinjan, 2014; Hormozdiari *et al.*, 2016; Krijger and De Laat, 2016; Yao *et al.*, 2017). Colocalisation can also highlight potential disease-relevant tissue types; evidence of colocalisation in only certain tissues can indicate that they may be of particular importance in disease development (Hormozdiari *et al.*, 2016). However, evidence of colocalisation in a certain tissue should not be seen as final evidence either as the vast majority of expression quantitative trait loci (QTLs) affect gene expression in several tissues (Ongen *et al.*, 2017; Wang *et al.*, 2019). Still, given the costs and time involved in functional follow-up of GWAS loci, a major benefit of colocalisation methods is that they can highlight a more limited set of genes and tissues with a higher likelihood of mediating disease risk for functional follow-up.

1.4.3.3. Coloc

One of the most commonly used colocalisation methods is Coloc (Giambartolomei *et al.*, 2014; Guo *et al.*, 2015; Franceschini *et al.*, 2018). Coloc only requires summary-level data and is

available as an R-package, which likely has contributed to its popularity (Giambartolomei *et al.*, 2014). The method relies on Bayesian statistics and requires setting prior probabilities for each of the traits having a causal variant (p1 and p2), as well as setting the prior probability of colocalisation (p12). The standard settings for these prior probabilities in the Coloc R package are $p1=p2=1\times10^{-4}$ and $p12=1\times10^{-5}$ (Giambartolomei *et al.*, 2014). A later simulation study by Wallace (2020), one of the Coloc originators, has indicated that while the standard settings for p1 and p2 are reasonable, $p12=1\times10^{-5}$ is likely too liberal. Wallace instead recommended a prior of $p12=5\times10^{-6}$ to assure robustness (Wallace, 2020).

Coloc evaluates the posterior probability (PP) for each of five hypotheses, given as a numeric value between 0 and 1 (Giambartolomei *et al.*, 2014):

a) H0: There is no causal variant for either trait.
b) H1: There is a causal variant for the first trait but not the second trait.
c) H2: There is a causal variant for the second trait but not the first trait.
d) H3: There are causal variants for both traits, but it is not the same genetic variant.
e) H4: There are causal variants for both traits, and it is the same genetic variant.

Of these, H3 corresponds to a scenario with evidence against colocalisation and H4 to a scenario with evidence in favour of colocalisation (Giambartolomei *et al.*, 2014). Different thresholds have been used to define colocalisation and to assess if there is enough power to determine colocalisation (the latter commonly computed by summing the PPs for H3 and H4) (Guo *et al.*, 2015; Franceschini *et al.*, 2018; Çaliskan *et al.*, 2019; Raulerson *et al.*, 2019).

In addition to the use of different priors, PP thresholds, and power thresholds, studies have also used different region sizes to apply Coloc on, commonly 400 kb and 1 Mb region sizes (Guo *et al.*, 2015; Franceschini *et al.*, 2018; Çaliskan *et al.*, 2019; Raulerson *et al.*, 2019). The choice of region size will effectively determine the number of SNPs included in the analysis, which also has been shown to affect the PP of colocalisation (Guo *et al.*, 2015; Wallace, 2020). Still, another research

group found Coloc to perform quite well in comparison to other methods, including methods using individual-level data (Chun *et al.*, 2017).

A major limitation of Coloc has been that it assumed a single causal variant per locus, and the method used to only assess the strongest signal (Giambartolomei *et al.*, 2014). The authors noted that a way to overcome this limitation could be to apply Coloc to summary statistics conditioned on each lead variant (Giambartolomei *et al.*, 2014). However, this may not always be possible in practice, for example due to missing data on effect allele frequencies or computational reasons (Yang *et al.*, 2012). However, in 2020 and 2021 updated versions of Coloc was released by Wallace (2020, 2021), which promises improved accuracy in determining colocalisation in the case of allelic heterogeneity.

1.4.3.4. eCAVIAR

Another commonly used colocalisation software is eCAVIAR (Hormozdiari *et al.*, 2016). Similar to Coloc, eCAVIAR only requires summary-level data for the genetic associations, albeit with individual-level data for an LD-reference panel with similar ancestral background (Giambartolomei *et al.*, 2014; Hormozdiari *et al.*, 2016). To quantify the evidence of colocalisation, eCAVIAR outputs a colocalisation posterior probability (CLPP) score for each genetic variant (Hormozdiari *et al.*, 2016). The CLPP score is a numerical value between 0 and 1, and corresponds to the posterior probability of the variant being causal for both traits (Hormozdiari *et al.*, 2016). A major strength with eCAVIAR is that it does not assume a single causal variant per locus, but only requires setting the maximum number of causal variants to be evaluated (Hormozdiari *et al.*, 2016). Yet, similar to Coloc, the software settings for eCAVIAR commonly vary between studies. For example, previous studies have included SNPs within various distance from the top disease-associated SNP (by P-value), from +/- 50 SNPs to all SNPs within a region of +/- 500 kb (Hormozdiari *et al.*, 2016; Aylward *et al.*, 2018; B. Liu, Calton, *et al.*, 2018; B. Liu, Pjanic, *et al.*, 2018)). Different thresholds have also been used to define colocalisation; CLPP-score thresholds ranging from 0.01 to 0.5 have been used in previous studies (Hormozdiari *et al.*,

2016; B. Liu, Calton, *et al.*, 2018; B. Liu, Pjanic, *et al.*, 2018; Lee *et al.*, 2018; Sakornsakolpat *et al.*, 2019). Finally, studies have also used different settings for the maximum number of causal variants (Hormozdiari *et al.*, 2016; B. Liu, Pjanic, *et al.*, 2018; Lee *et al.*, 2018; Zhang *et al.*, 2018).

1.4.3.5. Other colocalisation methods

There are several other colocalisation methods, including JLIM (Chun *et al.*, 2017), enloc (Wen, Pique-Regi and Luca, 2017), PICCOLO (Guo *et al.*, 2019), and HyPrColoc (Foley *et al.*, 2021). Each of them has their own strengths and limitations. For example, JLIM may outperform Coloc, but requires individual-level data (Giambartolomei *et al.*, 2014; Chun *et al.*, 2017). Enloc seems to give concordant but less conservative results as eCAVIAR (Hormozdiari *et al.*, 2016; Wen, Pique-Regi and Luca, 2017). However, enloc requires extensive pre-processing formatting that uses several different executables (Wen, 2019; Pividori *et al.*, 2020). An upgraded software (fastENLOC) was therefore released in 2020, which aimed to be both faster and easier to use (Wen, 2019; Pividori *et al.*, 2020).

PICCOLO and HyPrColoc are both based on the Coloc software (Giambartolomei *et al.*, 2014; Guo *et al.*, 2019; Foley *et al.*, 2021). The benefit of PICCOLO is that it does not require full summary statistics for the locus to evaluate colocalisation (Guo *et al.*, 2019). HyPrColoc, on the other hand, allows for assessment of colocalisation across more than two traits (Foley *et al.*, 2021).

1.5. Thesis aims

There are numerous examples of sexual dimorphism and sex differences in the clinical presentation of disease (DeVon *et al.*, 2008), risk factor-disease relationships (Karakas *et al.*, 2018; Ruth *et al.*, 2020), reference ranges of biomarkers (Säfström, Lindahl and Swahn, 2000; Mendis *et al.*, 2011; Kimenai *et al.*, 2018; Sobhani *et al.*, 2018; Apple *et al.*, 2020; National Institute for Health and Care Excellence, 2020), and effects of drugs (The Steering Committee of the Physicians' Health Study Research, 1988; United States General Accounting Office, 2001a; Marshall, 2005; Ridker *et*

al., 2005). Yet, this introduction has highlighted how sex-specific effects are underinvestigated and underreported, in study types ranging from clinical trials to animal studies to GWASs and MR studies (Beery and Zucker, 2011; Nelson *et al.*, 2017; GEFOS Consortium, 2018; Geller *et al.*, 2018; Mahajan, Taliun, *et al.*, 2018; CARDIoGRAMplusC4D Consortium, 2020; DIAGRAM, 2020). If we do not know when effects differ between men and women, we risk suboptimal prevention and treatment of diseases in both sexes.

I have chosen to explore sex differences in the causes and consequences of obesity, using genetic epidemiology, in my thesis. Obesity has wide-spread consequences on several of the leading causes of death (Prospective Studies Collaboration, 2009; GBD 2015 Obesity Collaborators *et al.*, 2017; The World Health Organization, 2018b, 2020b), yet sex-specific effects have not been systematically investigated using genetic epidemiology methods. Similarly, there is a lack of large-scale sex-stratified studies assessing the effects of proteins – commonly used as biomarkers and drug targets – on obesity and cardiometabolic diseases using genetic data. Finally, the obesity-related female reproductive diseases PCOS and miscarriage have been neglected in both research and the clinic, in spite of far-reaching consequences on patient health (Aleman *et al.*, 2005; Meyer, McGrath and Teede, 2005; Trinder *et al.*, 2006; Cumming *et al.*, 2007; Jones *et al.*, 2008; Broder-Fingert *et al.*, 2009; Moran *et al.*, 2010; Sivayoganathan *et al.*, 2011; Cantwell *et al.*, 2011; Fauser *et al.*, 2012; Joham *et al.*, 2015; Maino *et al.*, 2016; Qiao *et al.*, 2016; Chen, Fu and Huang, 2016; Pereza *et al.*, 2017; Gibson-Helm *et al.*, 2017; Wahabi *et al.*, 2018; Weng *et al.*, 2018; Wolf *et al.*, 2018; Lim *et al.*, 2019; Damone *et al.*, 2019; Greenwood *et al.*, 2019).

My thesis will therefore focus on three major aims;

I. Determine the effect of obesity in leading mortality causes in men versus women (Chapter 2).
II. Discover potential sex-specific protein biomarkers and drug targets for obesity and leading cardiometabolic diseases (Chapter 3).

III. Establish the effect of obesity on two obesity-associated female reproductive diseases and identify potential disease-mediating genes (Chapter 4).

Chapter 2. Sex-specific effects of obesity on the leading causes of death

2.1. Introduction

2.1.1. Previous studies of sex-specific effects of obesity on leading causes of death have been inconclusive

2.1.1.1. Differences in mortality between men and women for leading causes of death

As outlined in Chapter 1, it is clear that disease age-of-onset and mortality rates for common diseases often differ between the sexes (Fig 1.1) (Khan *et al.*, 2018; The World Health Organization, 2018a). For example, although approximately the same number of men and women die of stroke each year, men tend to have their first stroke at a younger age (Persky, Turtzo and McCullough, 2010; The World Health Organization, 2018a). It is also clear that sex differences exist in body morphology and fat distribution (Schreiner *et al.*, 1996; Demerath *et al.*, 2007). Observational studies have indicated that obesity may impact risk of cardiovascular disease differently in men versus women (Dagenais *et al.*, 2005; Li *et al.*, 2006; Hu *et al.*, 2007; Abete *et al.*, 2015; Dhana *et al.*, 2016; Rost *et al.*, 2018). Yet, previous observational studies have been contradictory and inconclusive, which can be exemplified with studies of sex-specific effects of WHR on risk of cardiovascular death (Dagenais *et al.*, 2005; Dhana *et al.*, 2016; Rost *et al.*, 2018).

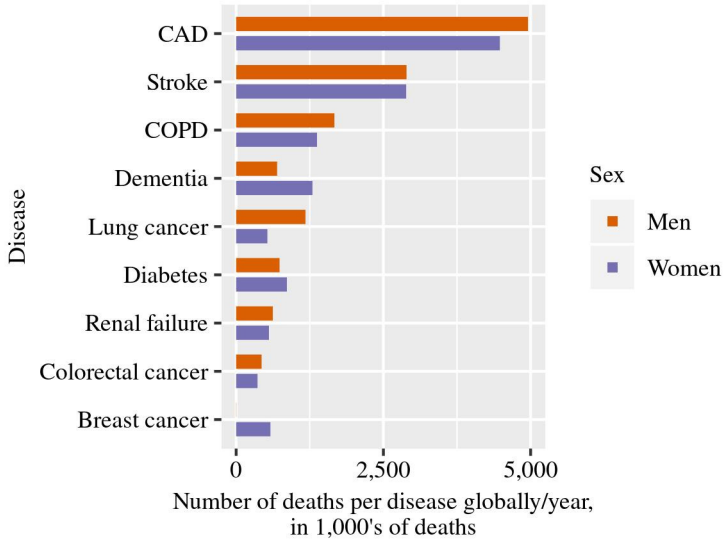


Figure 1.1. Sex-specific global mortality rates for the leading causes of death from non-communicable diseases, globally and in high-income countries. Estimates from the World Health Organization for 2016 (The World Health Organization, 2018a). COPD, chronic obstructive pulmonary disease.

2.1.1.2. An example of the inconclusive results from previous studies investigating sex-specific effects of obesity on common causes of death

A study by Rost *et al.* (2018) found WHR to be associated with cardiovascular death in both sexes, and a particularly good predictor of cardiovascular death in women (Table 2.1). On the contrary, Dhana *et al.* (2016), could not identify an effect of WHR on cardiovascular death in either sex. Yet another study, by Dagenais *et al.* (2005), did not find WHR to be associated with risks of cardiovascular death in men – but some evidence of an effect in women. Besides differences in study design, sample size and outcome definitions between the studies, a potential explanation could also be differences in the study populations, such as age and obesity levels. For example, Li *et al.* (2006) found evidence of WHR being associated with cardiovascular disease risk in women regardless of obesity level. However, there was only evidence of an association between WHR and cardiovascular disease in men with a normal weight (BMI<25 kg/m^2) (Li *et al.*, 2006).

Table 2.1. Examples of the contradictory results from observational studies investigating if WHR affects cardiovascular disease.

Study	Exposure	Outcome	N cases/total N men	N cases/total N women	Association in men	Association in women
Dagenais *et al.* (2005)	WHR	Death from cardiovascular disease	519/6620	136/2182	No	Yes
Dhana *et al.* (2016)	WHR	Death from cardiovascular disease	564/2,626	631/3,740	No	No
Li *et al.* (2006)	WHR	Cardiovascular disease events	816/10,369	464/16,638	Only in men with BMI<25 kg/m^2	Yes
Rost *et al.* (2018)	WHR	Death from cardiovascular disease	647/6,670	385/6,637	Yes	Yes

All studies identified a null or positive association between WHR and cardiovascular disease in both sexes.

Another possible explanation of the contradictory evidence from observational studies is study-differential bias from confounding and/or reverse causation (Burgess and Thompson, 2015). Such bias could potentially be minimised by using an MR study design (Burgess and Thompson, 2015). However, few MR studies have investigated the effect of abdominal fat on risk of stroke, and to my knowledge no studies have been performed sex-stratified. Even so, the sex-combined MR studies have also been inconclusive (Dale *et al.*, 2017; Emdin *et al.*, 2017). For example, Emdin *et al.* (2017) used a GRS of WHRadjBMI, consisting of 48 SNPs, to assess the effect of higher WHRadjBMI on overall stroke, but could not find any evidence of a causal effect. On the contrary, Dale *et al.* (2017) found suggestive evidence of higher WHRadjBMI increasing risk of ischaemic stroke.

2.1.2. The importance of teasing out sex-specific consequences of obesity

To date, few studies have investigated sex-specific consequences of obesity on common diseases using genetic epidemiology (T. Huang *et al.*, 2015; Lotta *et al.*, 2018). Huang *et al.* (2015) used 14 SNPs to investigate how higher WHR impacted on risk of T2D risk in men and women, and found that central obesity conferred comparable risks in both sexes. Lotta *et al.* (2018) used up to 202 genetic variants to create GRSs for both WHR as well as specific GRSs for gluteofemoral- and

abdominal fat distribution. Lotta *et al*. (2018) then used an interaction term for sex to assess if these fat distribution traits were sex-differentially associated with increased risks of CAD and T2D, but could not identify any sex differences.

In summary, although observational studies have indicated that obesity may confer different risks of disease in men and women, they have been contradictory (Dagenais *et al.*, 2005; Li *et al.*, 2006; Hu *et al.*, 2007; Abete *et al.*, 2015; Dhana *et al.*, 2016; Rost *et al.*, 2018). It is possible that confounding and/or reverse causation have impacted these previous studies (Burgess and Thompson, 2015). Yet, there have been few previous studies investigating the sex-specific consequences of obesity using an MR design (T. Huang *et al.*, 2015; Lotta *et al.*, 2018). These previous MR studies have used a limited number of SNPs and/or analyses to investigate sex differences for a limited number of outcomes (T. Huang *et al.*, 2015; Lotta *et al.*, 2018). There has been no systematic analysis of the sex-specific effects of obesity on leading causes of death using MR, nor have any previous studies systematically evaluated different approaches in constructing sex-specific GRSs. We need to know the sex-specific effects of obesity on leading diseases if we are to optimise prevention in each sex and understand sex-specific mechanistic pathways driving disease.

2.1.4. Aims

The main research aim for this chapter was to determine the effects of obesity on leading mortality causes in men versus women (Fig 2.2). To explore this question further, my research objectives were:

I. To evaluate different approaches to design sex-specific GRSs of obesity.
II. To establish the sex-specific effects of obesity on leading causes of death from non-communicable diseases using MR.

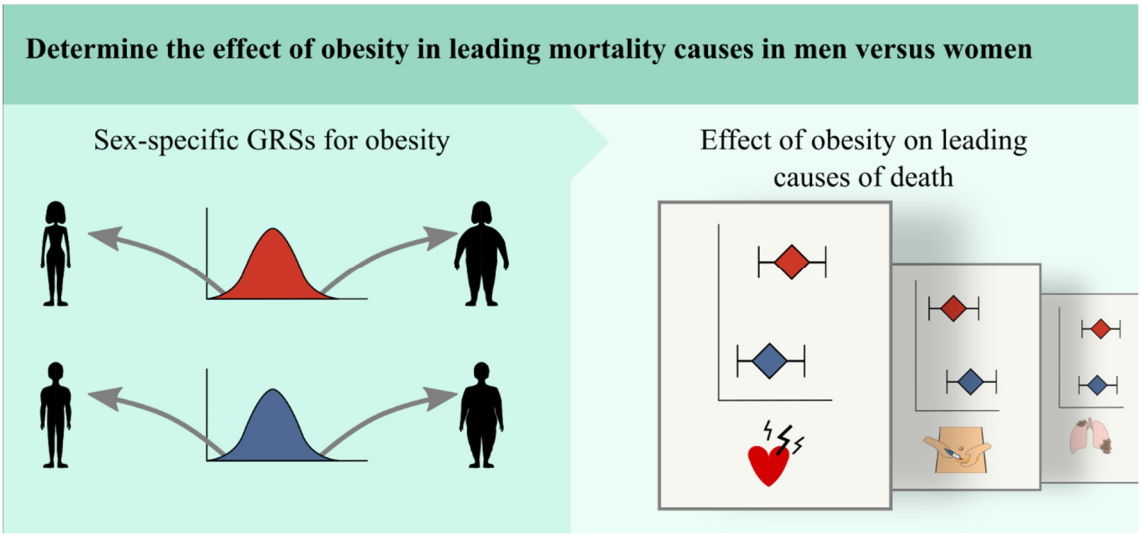


Figure 2.2. Graphical abstract of the primary research aim and research objectives for Chapter 2. The main research aim of Chapter 2 was to determine the effects of obesity on the leading causes of death in men versus women. To achieve this aim, I had two research objectives. First, to create and evaluate different approaches to construct sex-specific GRSs, and second, to establish the sex-specific effects of obesity on the leading causes of death from non-communicable diseases using MR.

2.2. Methods

2.2.1. Study overview

I first constructed and evaluated different approaches of creating sex-specific GRSs for BMI, WHR, and WHRadjBMI (Fig 2.3) (Burgess and Thompson, 2015). Analyses were performed using BMI, WHR, and WHRadjBMI as measurements of obesity and fat distribution, as they are easy to measure, are commonly used in the clinic, and have publicly available sex-stratified summary statistics from large-scale GWASs (The World Health Organization, 2000b, 2008; Pulit *et al.*, 2019). Thereafter, I investigated the association between the sex-specific instruments and leading causes of death using logistic regression (Osborne, 2017b; The World Health Organization, 2018b). GRS-disease combinations that surpassed the multiple testing threshold were brought forward for MR analysis and computation of formal MR estimates (Bonferroni, 1936; Davey Smith and Ebrahim, 2003). Similarly, I investigated associations between the sex-specific GRSs for obesity and cardiometabolic risk factors using a combination of linear regression, logistic regression, and MR (Davey Smith and Ebrahim, 2003; Osborne, 2017b, 2017a, 2017c). For all

analyses, sex differences were assessed using a Cochran's Q test (Cochran, 1954). Finally, I followed up with a range of sensitivity analyses to assure the robustness of my findings.

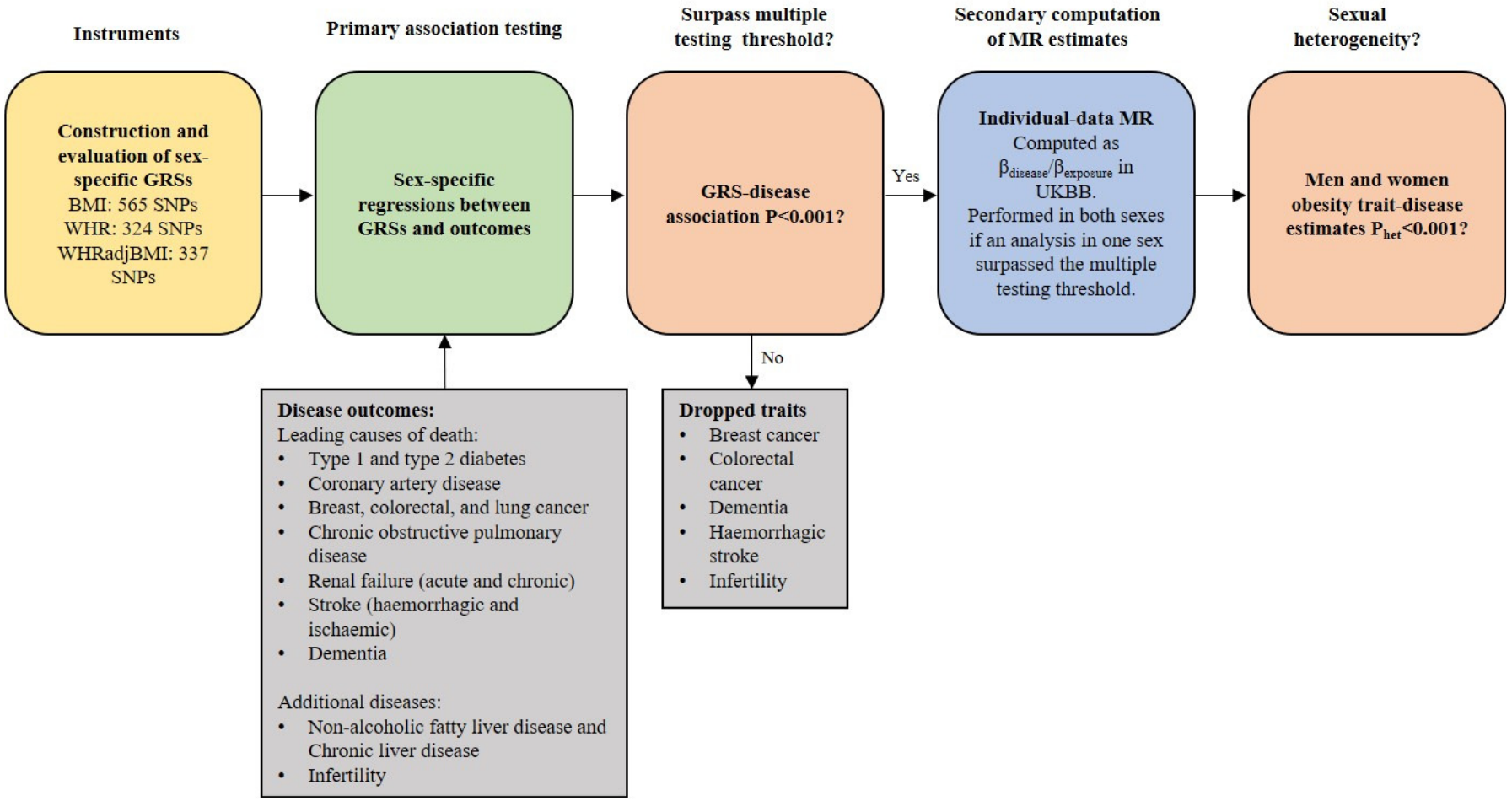


Figure 2.3. Study overview. First, sex-specific instruments for obesity were created (Burgess and Thompson, 2015). These were then investigated for an association with disease outcomes using logistic regression (Osborne, 2017b). Obesity trait-disease outcomes that surpassed the multiple testing threshold (P<0.001) in at least one sex were taken forward for formal computation of MR estimates (Bonferroni, 1936; Davey Smith and Ebrahim, 2003; Burgess and Thompson, 2015). Thereafter, sexual heterogeneity between the sex-specific estimates were assessed using a Cochran's Q test (Cochran, 1954). Figure adapted (changed text, streamlined flowchart, and updated layout) from Censin, Peters, et al. (2019) and the corresponding pre-print (Censin, Bovijn, et al., 2019). CAD, coronary artery disease; CLD, chronic liver disease; COPD, chronic obstructive pulmonary disease; NAFLD, non-alcoholic fatty liver disease;

T1D, type 1 diabetes; T2D, type 2 diabetes.

2.2.2. The UK biobank

2.2.2.1. Overview of the UK biobank

The UKBB is a prospective cohort study of over 500,000 individuals from the UK, with individual-level data available to researchers who meet the criteria for access (Sudlow *et al.*, 2015; Bycroft *et al.*, 2018). Participants were between 40 to 69 years of age when recruited in 2007 to 2010, and underwent extensive profiling at baseline, including anthropometric measurements, questionnaires, nurse interviews, and sample collection, across 22 different centres spread across the UK (Sudlow *et al.*, 2015; Bycroft *et al.*, 2018). Repeat assessments have been completed for a subset of the cohort (approximately 20,000 participants), and the biobank is regularly updated with electronic health records (Sudlow *et al.*, 2015; Bycroft *et al.*, 2018). The UKBB has a Research Tissue Bank approval, all participants gave informed consent, and this work was performed under UKBB study application ID 11867 (Sudlow *et al.*, 2015; Bycroft *et al.*, 2018).

2.2.2.2. DNA extraction and genotyping

Blood samples were collected at the baseline assessment (Sudlow *et al.*, 2015; Bycroft *et al.*, 2018). Most participants were genotyped using the customised UK Biobank Axiom array from Affymetrix, but a subset of 50,000 participants – selected based on lung function and/or smoking habits – were genotyped on the UK BiLEVE Axiom array (Bycroft *et al.*, 2018). The two arrays share approximately 95% of the assayed genetic variants (Bycroft *et al.*, 2018).

2.2.2.3. Pre-imputation quality control

Pre-imputation quality control was performed by the UKBB (Bycroft *et al.*, 2018). Genetic markers were excluded based on batch effects, plate effects, genotyping array effects, sex effects, Hardy-Weinberg equilibrium (HWE) $P<1\times10^{-22}$, and disagreement between control duplicates (Bycroft *et al.*, 2018). Principal components (PCs) for ancestry were computed using fast PC

analysis (fastPCA) (Galinsky *et al.*, 2016) using markers passing quality control in unrelated samples (no samples related to a third degree or higher) (Bycroft *et al.*, 2018).

2.2.2.4. Sample quality control

The UKBB had identified samples with poor quality based on a set of high quality genetic markers – for a detailed description see Bycroft *et al*. (2018). Apart from exclusion of duplicated samples, the UKBB (Bycroft *et al.*, 2018) also provided information of the samples with:

a) Withdrawn consent.
b) Excess heterozygosity, adjusted for population structure.
c) More than 5% missing markers.
d) Mis-match between self-reported and genetically inferred sex.
e) Suspected sex chromosome aneuploidy.
f) More than 10 third degree relatives.
g) Samples that were excluded in kinship calculations or autosome phasing because of, for example, high degrees of missingness and/or being heterozygosity outliers.

I used the data provided by the UKBB for sample quality control and excluded all samples failing any of the criteria above. Kinship coefficients had been computed by the UKBB using Kinship-based INference for Genome-wide association studies (KING) (Manichaikul *et al.*, 2010). I excluded one individual in each pair estimated to be related to a second degree or higher (kinship coefficient>0.0884) (Manichaikul *et al.*, 2010; Bycroft *et al.*, 2018). Analyses were primarily conducted in Europeans, but also in the curated "White British" ancestry subset provided by the UKBB, consisting of individuals that self-reported as "British" and who had a similar background based on principal components analysis (Bycroft *et al.*, 2018; Rutten-Jacobs *et al.*, 2018; Celis-Morales *et al.*, 2019).

I further excluded individuals that reported incompatible ancestry codes in different assessments (for example, self-reported as "White" in the baseline assessment and any of the non-"White" codes in another assessment, i.e., codes starting with and without "1" in data field 21000,

respectively). LD between SNPs was computed in participants passing quality control after further pruning the sample to contain no individuals related to a third degree or higher (kinship coefficient>0.0442, N=380,262) (Manichaikul *et al.*, 2010; Bycroft *et al.*, 2018).

2.2.2.5. Post-imputation quality control

The UKBB provided genotype data in BGEN v1.2 format, which I converted into hard calls (--hard-call-threshold 0.1) using PLINK v2.00aLM (Purcell *et al.*, 2007). I then performed post-imputation quality control using PLINK v1.90b3 (Purcell *et al.*, 2007) and excluded SNPs with HWE $P<1\times10^{-6}$, missing call rate >0.05, minor allele frequency <0.01%, and imputation info score ≤0.3, as suggested by the UKBB and/or used previously in large genetic studies of the UKBB (Bycroft *et al.*, 2018; Emdin *et al.*, 2018; Yengo *et al.*, 2018; Morris *et al.*, 2019; Pulit *et al.*, 2019).

2.2.3. Obesity traits

I used baseline measurements for BMI, WHR (computed as the waist circumference divided by the hip circumference), and for deriving the WHRadjBMI measurements. In the main analyses, I standardised these traits through rank-inverse normal transformation (RINT) (Bliss, 1967) of the residuals after regression of the trait on age and age^2 at baseline, assessment centre, as well as sex for the sex-combined analyses (Emdin *et al.*, 2017; Pulit *et al.*, 2019). This was done separately in the sex-specific analyses, but jointly in the sex-combined analyses, after any sample quality exclusions. For WHRadjBMI, I also adjusted for BMI (Pulit *et al.*, 2019). For analyses involving anthropometric and/or blood pressure phenotypes, I also excluded participants whose measurements were likely to be temporarily affected by special factors; a) up to 29 participants with BMI <15kg/m^2 (to minimise the risk of including participants whose low BMI was caused by severe disease) and b) up to 279 women who reported that they were pregnant or unsure about pregnancy. BMI in kg/m^2 and WHR as a ratio was used for the analyses done in clinical units.

2.2.4. Sex-specific genetic risk scores for obesity

2.2.4.1. Summary statistics for genetic variants included in the genetic risk scores

I used SNPs from a large meta-analysis of GIANT and UKBB (Pulit *et al.*, 2019), which had sex-specific and sex-combined summary statistics available for BMI, WHR and WHRadjBMI. The study had included up to 374,756 men and 434,794 women for BMI, 316,772 men and 381,152 women for WHR, and 315,284 men and 379,501 women for WHRadjBMI, all of European ancestry (Pulit *et al.*, 2019). Sex-specific primary ("index") variants associated with each obesity trait at $P<5\times10^{-9}$ were identified using clumping and proximal and joint conditional analysis (Yang *et al.*, 2011) by Pulit *et al.* (2019). Briefly, associated loci were identified through clumping of all SNPs associated with the obesity trait P<0.05 and ±5 Mb and in LD ($r^2>0.05$) with a top SNP (at $P<5\times10^{-9}$) for the obesity trait under investigation (Pulit *et al.*, 2019). A 1 kb buffer up- and downstream was added to each locus, before overlapping loci were merged (Pulit *et al.*, 2019). Primary and secondary variants $P<5\times10^{-9}$ were then identified by Pulit *et al.* (2019) using the --cojo-slct function in Genome-wide Complex Trait Analysis (GCTA) (Yang *et al.*, 2011).

In the GWAS by Pulit *et al.* (2019), sex differences in the men and women effect estimates had been evaluated for each SNP by computing $P_{sexdiff}$-values as:

$$t_{diff} = \frac{\beta_{men} - \beta_{women}}{\sqrt{SE^2_{men} + SE^2_{women} - 2r \cdot SE^2_{men} \cdot SE^2_{women}}}$$

with β equal to the SNP estimate and *SE* the SNP standard error in men and women, respectively, and *r* the overall Spearman rank correlation coefficient for the SNP effects between men and women (Randall *et al.*, 2013; Pulit *et al.*, 2019).

2.2.4.2. Construction of sex-specific genetic risk scores

I constructed and evaluated seven different approaches for making sex-specific GRSs for BMI, WHR, and WHRadjBMI, with the aim to create GRSs that potentially could (Fig 2.4):

a) Take advantage of the possibly more exact *sex-combined* effect estimates for *non-sex-differential* SNPs.

b) Take advantage of the possibly more exact *sex-specific* effect estimates for the *sex-differential* SNPs.

For all approaches except one, the same SNPs were included for both the men, women and sex-combined instruments. These instruments included the primary index SNPs $P<5\times10^{-9}$ identified in the men, women, or sex-combined GWAS of each obesity trait (Pulit *et al.*, 2019). I then excluded the SNP with the highest sex-combined P-value within each 1 Mb sliding window to avoid correlations between SNPs (see yellow box in Fig 2.4). Thereafter, I weighted the SNPs included in each instrument using different approaches:

Approach 1-2. Combined and **sex-specific estimates approaches:** These two approaches use the same SNPs for the men, women, and sex-combined analyses, but weights each SNP by either their sex-combined effect estimates or their sex-specific effect estimates, respectively. The combined estimates approach was also applied in men- and women-stratified analyses for comparison.

Approach 3-6. P-Sexdiff Bonferroni and **P-Sexdiff false discovery rate (FDR) approaches:** These four approaches also used the same SNPs for the men and women-stratified analyses. SNPs were weighted by the sex-specific estimate if the SNP $P_{sexdiff}$-value was below the specified sex difference threshold, and by the sex-combined estimate if the SNP $P_{sexdiff}$-value was above the sex difference threshold (Randall *et al.*, 2013; Pulit *et al.*, 2019). Four sex difference thresholds were evaluated: Bonferroni-threshold (0.05/number of SNPs) (Bonferroni, 1936), 1% FDR threshold, 5% FDR threshold, and 10% FDR threshold (all thresholds computed in each instrument) (Benjamini and Hochberg, 1995), resulting in four P-Sexdiff approaches.

Approach 7. Primary SNPs in each sex approach: This approach only included the primary SNPs $P<5\times10^{-9}$ identified in each sex (Pulit *et al.*, 2019). All SNPs were then weighted by their sex-specific estimate.

I then excluded six SNPs found to be non-biallelic in the UKBB (rs805768, rs699370, rs10892873, rs10828247, rs73646205, and rs7798002), as these could invalidate the additive assumptions of the GRSs (see white box in Fig 2.4). I further excluded four SNPs that had failed my post-imputation

quality control in the UKBB (rs11500477, rs144926207, rs547943994, and rs9638713). Finally, I computed long-distance LD using PLINK v1.90b3 (Purcell *et al.*, 2007) between all pairs of SNPs included in an instrument, and excluded the SNP with the highest P-value in all pairs of SNPs correlated $r^2>0.05$ (rs8189480, rs9302652, and rs1025395). In all analyses, I aligned SNPs so that the effect allele (EA) corresponded to a higher level of the obesity trait under investigation. I then computed a summed GRSs for each participant, weighted by the effect estimates, using the "--score" option in PLINK v1.90b3 (Purcell *et al.*, 2007).

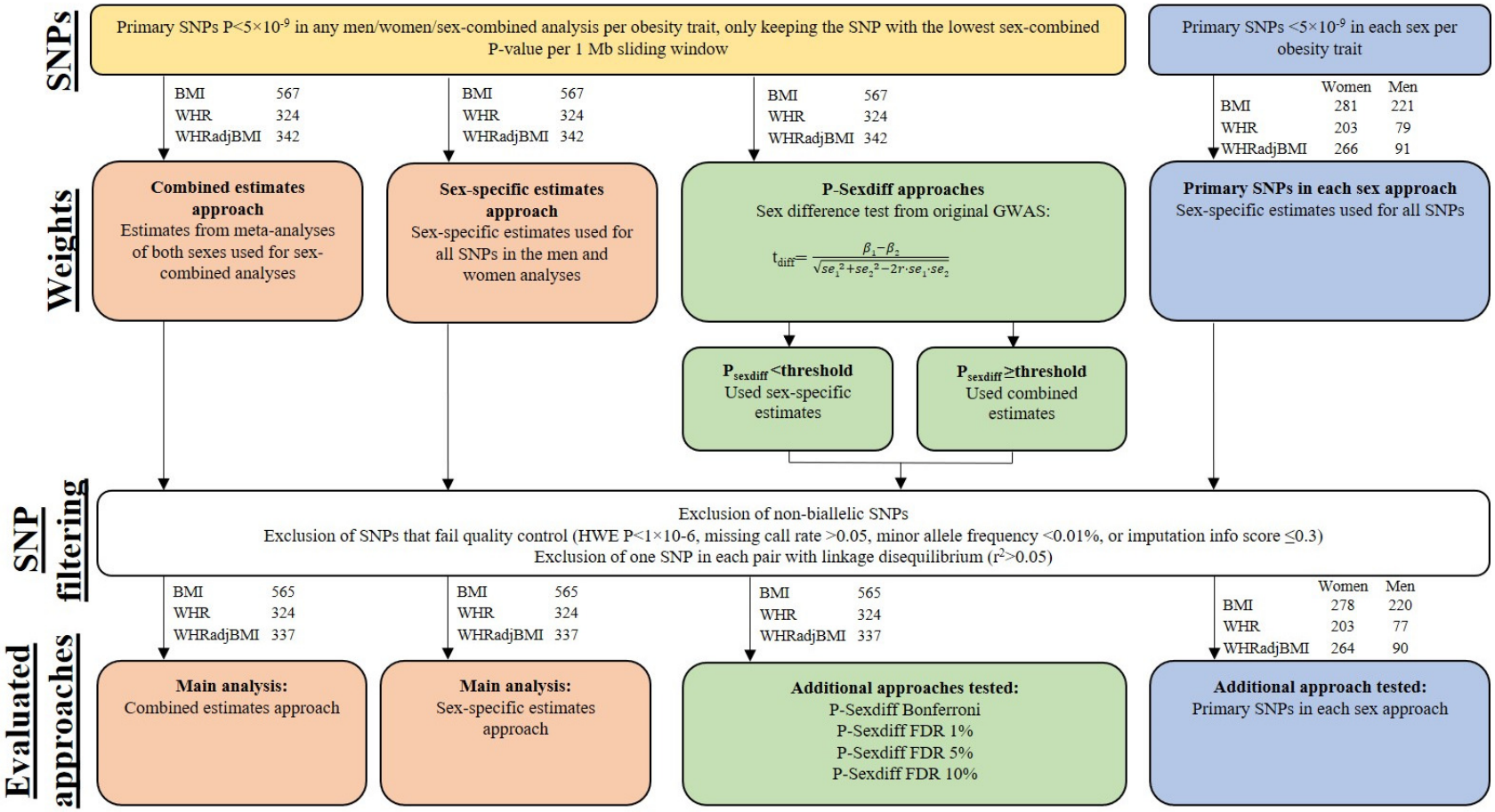


Figure 2.4. Instrument selection and weighting flowchart. Seven different approaches to construct sex-specific GRSs for obesity were created and evaluated. SNPs and weights were selected from a meta-analysis of GIANT and the UKBB (Pulit *et al.*, 2019). The main sex-specific GRS approach in coral, the other five other approaches that were evaluated in green and blue. For the main approaches as well as for the P-Sexdiff approaches, the same set of primary SNPs associated with each obesity trait ($P<5\times10^{-9}$) in men, women, or sex-combined were included. For the Primary SNPs in each sex approach,

only SNPs $P<5\times10^{-9}$ in the investigated sex were included (Pulit *et al.*, 2019). SNPs were then further filtered to assure quality and independence, and weighted as described. Figure adapted (updated layout and changed text) from Censin, Peters, et al. (2019) and the corresponding pre-print (Censin, Bovijn, et al., 2019).

2.2.4.3. Evaluation of sex-specific genetic risk scores

To assess the strengths of the instruments, I performed linear regression of each sex-specific GRS and their corresponding sex-standardised obesity trait (see section 2.2.3), adjusting for 10 PCs and array type (Emdin *et al.*, 2017; Osborne, 2017a, 2017c). I computed F-statistics using the formula:

$$F = \left(\frac{n-k-1}{k}\right)\left(\frac{R^2}{1-R^2}\right)$$

where n is the number of samples, $k = 1$ for the instrument, and R^2 the trait variance explained (Burgess and Thompson, 2011). Trait variance explained was computed as the increase in R^2 when including the GRS in the model (Osborne, 2017c). Sexual heterogeneity between the men- and women-specific estimates computed using each instrument creation approach was assessed using Cochran's Q test (Cochran, 1954). The GRS approach with the highest trait variance explained and lowest sexual heterogeneity in the association with the obesity trait was selected for use in downstream analyses.

2.2.5. Disease case and control criteria

I analysed all non-communicable diseases among the top ten leading causes of death (globally and in high-income countries) in 2016, as estimated by the World Health Organization (The World Health Organization, 2018b). This included CAD, stroke, COPD, dementia, lung cancer, diabetes, colorectal cancer, renal failure and breast cancer (The World Health Organization, 2018b). For stroke, I investigated ischaemic, haemorrhagic, and stroke of any cause. Similarly, I assessed acute renal failure, chronic renal failure, and renal failure of any cause. I analysed breast cancer in women only, since the vast majority of deaths from breast cancer occur in women (The World Health Organization, 2018a). I investigated T1D and T2D separately, in spite of the World Health Organization only providing mortality estimates for T1D and T2D combined, as these are considered to be separate diseases (The World Health Organization, 2018a, 2020a). In addition to these diseases, I also included NAFLD, CLD, and infertility in the study. These diseases represent increasing burdens of disease and have been previously linked to obesity (van der Steeg *et al.*,

2008; Macaluso *et al.*, 2010; Vernon, Baranova and Younossi, 2011; Williams *et al.*, 2011; Chalasani *et al.*, 2012; Campbell *et al.*, 2015; Estes *et al.*, 2018).

I decided on all case and control criteria in discussion with another licenced medical practitioner, Dr. Jonas Bovijn (Appendix Table 2.1). For CAD, we decided to use similar criteria as a large GWAS (Nelson *et al.*, 2017). For T1D and T2D, we decided to extract cases and controls based on a validated algorithm for prevalent T1D and T2D (Eastwood *et al.*, 2016). The algorithm outputs "probable" and "possible" cases for each diabetes type (Eastwood *et al.*, 2016). I included both probable and possible cases in the main analyses, but also performed sensitivity analyses restricting the cases to those being a "probable" case only. For controls, we decided to extract those the algorithm denoted as unlikely to have any kind of diabetes (Eastwood *et al.*, 2016).

Criteria for cases and controls for the other disease outcomes were defined using a combination of hospital outcome codes, data from the interview with a trained nurse in the UKBB assessments, and self-reported data (Appendix Table 2.1). For CAD, stroke (including ischaemic, haemorrhagic and of any cause), and acute and chronic renal failure, I also excluded samples from each control group after defining each case group. These exclusions were made to reduce the risk of including dubious cases as controls. For example, samples that reported having had a transient ischaemic attack were excluded from the stroke control group.

2.2.6. Cardiometabolic risk factors

I investigated five potential cardiometabolic risk factors: systolic blood pressure (SBP), diastolic blood pressure (DBP), fasting glucose (FG), fasting insulin (FI), and smoking status. These risk factors were selected as the World Health Organization (2017) has highlighted them as being important risk factors for cardiovascular disease, while also having reliable measurements available in the UKBB and/or sex-stratified GWAS summary statistics available at the time of the analysis (Sudlow *et al.*, 2015; Lagou *et al.*, 2021). For SBP and DBP, I used the mean values of the baseline measurements. To account for blood pressure medications, I added 15 mmHg to SBP and

10 mmHg to DBP if the participant had self-reported use of such (as reported in data-fields 6153 and 6177) (Tobin *et al.*, 2005; International Consortium for Blood Pressure Genome-Wide Association Studies *et al.*, 2011). In the main analyses, I then standardised SBP and DBP through RINT of the residuals after having regressed the trait on baseline age, age^2, assessment centre, and sex in the sex-combined analyses (Bliss, 1967; Pulit *et al.*, 2019). For some analyses, I also analysed associations with SBP and DBP in clinical mmHg units. Smoking status was based on self-report of being a previous or current smoker, or having smoked or currently smoking either most days or occasionally (code 1 or 2 in any of the data-fields 1239, 1249, or 20116).

My colleague, Dr. Reedik Mägi from the MAGIC consortium, kindly provided sex-specific summary-level data for serum FI (in pmol/L, ln-transformed) and plasma FG (in mmol/L, untransformed), now publicly available (Lagou *et al.*, 2021). For samples where FG had been measured in whole blood in the original GWAS, it had been corrected to plasma levels using a correction factor of 1.13 (Lagou *et al.*, 2021). SNP information (genomic position in GRCh 37 and rs-aliases) were updated as needed using the files "All_20150605.vcf.gz" (available at ftp://ftp.ncbi.nih.gov/snp/organisms/archive/human_9606_b144_GRCh37p13/VCF/) and RsMergeArch.bcp.gz" (available at ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606/database/organism_data/) from the National Center for Biotechnology Information (Sherry *et al.*, 2001).

2.2.7. Statistical analyses

2.2.7.1. Primary association testing between GRSs and disease outcomes

I first explored the associations of the sex-specific GRSs with the disease outcomes using logistic regression (Osborne, 2017b). GRS-disease combinations that surpassed multiple testing correction were then taken forward for quantification of the effect estimate using MR (Bonferroni, 1936). The significance threshold was set to $P<0.001$ for GRS-disease investigations (for 0.05/51 obesity trait-disease combinations) (Bonferroni, 1936). If a regression analysis supported an effect in either men

or women, MR was performed in both sexes to enable assessment of sexual heterogeneity (Cochran, 1954; Davey Smith and Ebrahim, 2003; Osborne, 2017b).

2.2.7.2. Mendelian randomisation analyses of obesity traits on disease outcomes

MR was performed using the ratio method, with the estimate being the ratio between the outcome and the obesity trait regressed separately on the obesity GRS (Wald, 1940; Burgess and Thompson, 2015). To account for the uncertainty in both regressions, I computed standard errors using the first two terms of the delta method (Bautista *et al.*, 2006; Thomas, Lawlor and Thompson, 2007; Burgess and Thompson, 2015). For MRs with a binary outcome, the GRS-obesity trait regression was performed in the controls for that disease to reduce the risk of reverse causation affecting the effect estimates (Burgess, Small and Thompson, 2017). In analyses involving traits that had been standardised through RINT of the residuals, I only adjusted for array and 10 PCs (as adjustment for age, age^2, assessment centre, and sex in the sex-combined analyses had been performed in the derivation of the residuals) (Bliss, 1967; Emdin *et al.*, 2017; Pulit *et al.*, 2019). For analyses conducted in clinical units, I adjusted for baseline age, age^2, assessment centre, array type, 10 PCs, as well as sex in the sex-combined analyses (Emdin *et al.*, 2017; Pulit *et al.*, 2019).

2.2.7.3. Mendelian randomisation analyses of obesity traits on cardiometabolic risk factors

I performed MR analysis of the obesity traits on the five cardiometabolic risk factors. For SBP, DBP, and smoking status, I computed MR estimates using the same methods as for the disease outcomes, i.e. using the ratio method to obtain an estimate and the delta method to compute standard errors (Wald, 1940; Bautista *et al.*, 2006; Thomas, Lawlor and Thompson, 2007; Burgess and Thompson, 2015). However, I only had access to summary-level data for FG and FI (Lagou *et al.*, 2021). I therefore used two-sample MR methods for these traits, including the IVW method, the MR-Egger method, and the weighted median method (Burgess, Butterworth and Thompson, 2013; Bowden, Davey Smith and Burgess, 2015; Bowden *et al.*, 2016; Yavorska and Burgess, 2017).

2.2.7.4. Assessing sexual heterogeneity

I assessed sexual heterogeneity using a Cochran's Q test (Cochran, 1954) between equivalent men and women analyses (including linear and logistic regressions, MRs, and relevant sensitivity analyses) (Davey Smith and Ebrahim, 2003; Bowden, Davey Smith and Burgess, 2015; Burgess *et al.*, 2015; Bowden *et al.*, 2016; Osborne, 2017a, 2017b, 2017c).

2.2.8. Sensitivity analyses

2.2.8.1. Mendelian randomisation using additional weighting strategies

To ascertain robustness, I reperformed the MR analyses using additional weighting strategies, in addition to the main approaches (approaches 1 and 2):

Approach 8. Unweighted approach: Weighting the SNPs included in the main approach by one.

Approach 9. Externally weighted approach: Weighting the SNPs used in the main approach using external datasets that did not include the UKBB cohort (Locke *et al.*, 2015; Shungin *et al.*, 2015). In total, 478 SNPs for were available for the BMI GRSs in all sex-strata, 264 SNPs in the sex-combined WHR GRS, 258 SNPs in the women WHR GRS, 253 SNPs in the men WHR GRS, 274 SNPs in the sex-combined WHRadjBMI GRS, 262 SNPs in the women WHRadjBMI GRS, and 247 SNPs in the men WHRadjBMI GRS.

The rationale for reperforming the MRs using an unweighted and externally weighted approach was to reduce the risk of overfitting and weak instrument bias affecting my results, as these methods have been shown to provide unbiased, but potentially less precise, estimates in simulation studies (Burgess and Thompson, 2013; Burgess, Dudbridge and Thompson, 2016).

More than twice as many index SNPs had been identified in women than in men for both WHR and WHRadjBMI in the original obesity GWAS (Fig 2.4, blue boxes) (Pulit *et al.*, 2019). I therefore also reconducted the MR analyses only including the primary SNPs identified in the sex-combined study ($P<5\times10^{-9}$) to reduce the risk of the higher number of primary SNPs identified in

women affecting my results (Pulit *et al.*, 2019). In a similar manner as for the other instruments, I excluded SNPs that failed genotype quality control (rs144926207) or that were non-biallelic in the UKBB (rs7798002, rs805768, rs10892873, rs805768, and rs699370). I also evaluated long-distance LD (r^2>0.05) using PLINK v1.90b3 (Purcell *et al.*, 2007) between all pairs of SNPs included in an instrument and excluded the genetic variant with the highest sex-combined P-value in each pair (rs9302652, rs7020604) (Pulit *et al.*, 2019). After exclusions, 669 genetic variants remained in the BMI instrument, 314 SNPs in the WHR instruments, and 341 SNPs for WHRadjBMI. These were then weighted using three different weighting strategies:

Approach 10. Sex-combined primary SNPs and combined weights approach: Weighting the SNPs by their sex-combined estimate.

Approach 11. Sex-combined primary SNPs and sex-specific weights approach: Weighting the SNPs by their sex-specific estimate.

Approach 12. Sex-combined primary SNPs and unweighted approach: Weighting the SNPs by one.

2.2.8.2. Assessing the impact of normalisation approach and units on sexual heterogeneity

The sex-specific MR estimates were computed as odds ratio (OR) per one *sex-specific* SD higher obesity trait. Thus, sex differences in the effect estimates can arise from sex-differential obesity trait distributions, where one SD corresponds to different amounts in the underlying clinical unit in men versus women. I therefore reconducted the main analyses with the obesity trait in clinical units, as well as normalising the obesity traits sex-combined (i.e., perform a RINT of the residuals from a linear regression including both men and women and adjusting for sex, in addition to the other adjustments) (Bliss, 1967). For the obesity trait-blood pressure MR associations, I also reconducted the analyses with blood pressure in clinical units (mmHg).

2.2.8.3. Analyses performed in British ancestry subset

Concerns have been raised about potential population stratification in both GIANT and the UKBB (Berg *et al.*, 2019; Haworth *et al.*, 2019). I therefore reanalysed the data restricting the sample to those in the curated set of "White British ancestry subset" that was provided by the UKBB

(Bycroft *et al.*, 2018). These individuals had both self-reported being "British" and had a similar ancestry based on outlier detection in the PC analysis (Bycroft *et al.*, 2018).

2.2.8.4. Analyses performed using the same number of cases and controls

For some outcomes, the number of cases differed substantially between men and women. For example, there were only 85 male infertility cases but 1,593 female infertility cases. As the power to detect associations would likely be very different between men and women in such cases, I reperformed the regression analyses in sub-cohorts that had been randomly subsampled to contain the same number of men and women cases and controls (Osborne, 2017b).

2.2.8.5. Analyses performed using stricter diabetes case criteria

I used a validated algorithm from Eastwood *et al.* (2016) for prevalent diabetes type status in the UKBB (Sudlow *et al.*, 2015). This algorithm defined T1D and T2D cases as either "possible" or "probable", with more evidence for the specific diabetes type in those defined as "probable" cases. I therefore also performed the MR analyses using "probable" cases only to minimise the risk misspecification of diabetes types affecting my results.

2.2.8.6. Pleiotropy-robust sensitivity analyses

I also reperformed the disease MRs using methods that give more consistent effect estimates in the case of pleiotropic instruments (Bowden, Davey Smith and Burgess, 2015; Bowden *et al.*, 2016). To obtain disease effect estimates for individual SNPs in the UKBB, I used logistic regression, adjusting for age, sex, array, and 10 PCs (Emdin *et al.*, 2017; Osborne, 2017b). For a limited number of SNPs, the model did not converge for some diseases and in such cases the SNP was excluded in downstream analyses.

I then conducted the MR-Egger intercept test and recomputed MR estimates using the MR-Egger and weighted-median methods using:

Approach 2. Sex-specific estimates approach: The approach used in the main analysis.

Approach 9. Externally weighted approach: The externally weighted approach using GIANT 2015 weights (Burgess, Butterworth and Thompson, 2013; Locke *et al.*, 2015; Shungin *et al.*, 2015; Bowden *et al.*, 2016; Yavorska and Burgess, 2017).

As these MR methods require information on standard errors for the instruments, I did not conduct them using the Unweighted approach (approach 8) (Bowden, Davey Smith and Burgess, 2015; Bowden *et al.*, 2016). I then assessed sexual heterogeneity between similar approaches and traits using Cochran's Q test (Cochran, 1954).

2.2.9. Data availability and software used

2.2.9.1. Data availability

For data availability see Appendix Table 1.1.

2.2.9.2. Software

I used PLINK v2.00aLM and PLINK v1.90b3 (Purcell *et al.*, 2007) for genotype data handling. I also used Python 3.5.2 (Python Software Fondation, 2020), R version 3.4.3 (R Core Team, 2014), bash version 4.1.2(2) and awk (Free Software Foundation, 2020) for data handling. The following programming packages were used; "pandas" (McKinney, 2010), "numpy" (Oliphant, 2006), "dplyr" (Wickham *et al.*, 2017), "MendelianRandomization" (Yavorska and Burgess, 2017), "ggpubr" (Kassambara, 2020), "grDevices" (R Core Team, 2014), "grid" (R Core Team, 2014), "lattice" (Sarkar, 2008), "gridExtra" (Auguie, 2017), "mada" (Doebler, 2017), and "ggplot2" (Wickham, 2016).

2.3. Results

2.3.1. Participant characteristics

A total of 195,041 men and 228,466 women were included in the analyses (Table 2.2). Mean BMI was higher in men (27.9 kg/m^2) than in women (27.0 kg/m^2), but women had a larger variation in

BMI, with 1 SD corresponding to 5.1 kg/m^2 in women but only 4.2 kg/m^2 in men. The male participants also had a higher mean WHR than the female participants (0.94 in men compared to 0.82 in women), but the trait variation was similar in both sexes (1 SD corresponding to a ratio of 0.07 in both men and women).

Table 2.2: Participant characteristics in the UKBB.

Characteristic	Men	Women
Individuals, N (%)	195,041 (46.1)	228,466 (53.9)
British, N (%)	174,261 (89.3)	201,958 (88.4)
Age, mean (SD), years	57.0 (8.1)	56.6 (7.9)
UK BiLEVE array, N (%)	23,216 (11.9)	22,848 (10.0)
Body mass index, mean (SD), kg/m2	27.9 (4.2)	27.0 (5.1)
Waist circumference, mean (SD), cm	97.1 (11.3)	84.5 (12.5)
Hip circumference, mean (SD), cm	103.5 (7.6)	103.3 (10.3)
Waist-hip-ratio, mean (SD)	0.94 (0.07)	0.82 (0.07)
Systolic blood pressure, mean (SD), mmHg	144.8 (19.4)	137.9 (21.2)
Diastolic blood pressure, mean (SD), mmHg	86.6 (11.0)	82.3 (11.1)
Breast cancer cases, N (%)	-	14,325 (6.3)
CAD, N (%)	24,444 (12.5)	11,578 (5.1)
CLD, N (%)	824 (0.4)	543 (0.2)
Colorectal cancer cases, N (%)	3,150 (1.6)	2,380 (1.0)
COPD cases, N (%)	7,874 (4.0)	6,799 (3.0)
Dementia cases, N (%)	583 (0.3)	447 (0.2)
Infertility cases, N (%)	85 (0.0)	1,593 (0.7)
Lung cancer cases, N (%)	1,477 (0.8)	1,239 (0.5)
NAFLD cases, N (%)	909 (0.5)	775 (0.3)
Renal failure cases, N (%)	5,697 (2.9)	3,909 (1.7)
Renal failure, acute, cases, N (%)	3,035 (1.6)	1,651 (0.7)
Renal failure, chronic, cases, N (%)	2,581 (1.3)	2,018 (0.9)
Stroke cases, N (%)	6,334 (3.2)	4,454 (1.9)
Stroke, haemorrhagic, cases, N (%)	928 (0.5)	979 (0.4)
Stroke, ischaemic, cases, N (%)	2,171 (1.1)	1,180 (0.5)
Type 1 diabetes cases, N (%)	826 (0.4)	677 (0.3)
Type 2 diabetes cases, N (%)	11,785 (6.0)	6,544 (2.9)

Table adapted (updated spelling and layout) from Censin, Peters, et al. (2019) and the corresponding pre-print (Censin, Bovijn, et al., 2019).

The number of cases also differed between the sexes for some of the investigated diseases. For example, there were more than twice as many male than female CAD cases and more than ten-fold as many female than male infertility cases.

2.3.2. Instrument strength

The F-statistics for instrument strength ranged from 2,842 (Primary SNPs in each sex approach for WHR in men) to 17,263 (Sex-specific estimates approach for WHRadjBMI in women) for the

evaluated GRSs in the sex-stratified analyses (Appendix Table 2.2) (Burgess and Thompson, 2011).

The GRSs had been computed as the sum of the obesity-increasing alleles, weighted by each allele's association with the obesity trait in SD-units. Thus, in theory, each 1-SD higher GRS would be expected to associate with 1 SD higher obesity trait in a linear regression (Osborne, 2017a, 2017c). However, the estimated increase in obesity per 1-SD higher GRS deviated from 1 for several of the GRSs (Fig 2.5). In the regressions between the GRSs for abdominal obesity and their respective trait, the female effect estimates tended to be >1 and the male effect estimates <1 for the approaches where some or all SNPs had been weighted by their sex-combined effect estimates. The instruments where SNPs had been weighted by sex-combined effect estimates also exhibited sexual heterogeneity in the linear regression estimates for abdominal obesity (P-value for sexual heterogeneity using Cochran's Q test (P_{het}) threshold <0.002, for 0.05/21 male-female estimate comparisons) (Bonferroni, 1936; Cochran, 1954). However, there was no significant sexual heterogeneity for any obesity-GRS when using the Sex-specific estimates approach. The Sex-specific estimates approach also had the highest trait variance explained in the sex-stratified analyses, ranging between 2.5-7.1%, and the highest F-statistics, ranging between 4,921–17,263 (Burgess and Thompson, 2011; Osborne, 2017c). I therefore decided to use the Sex-specific estimates approach as the main approach in the downstream analyses, with the corresponding Sex-combined approach used in the sex-combined analyses.

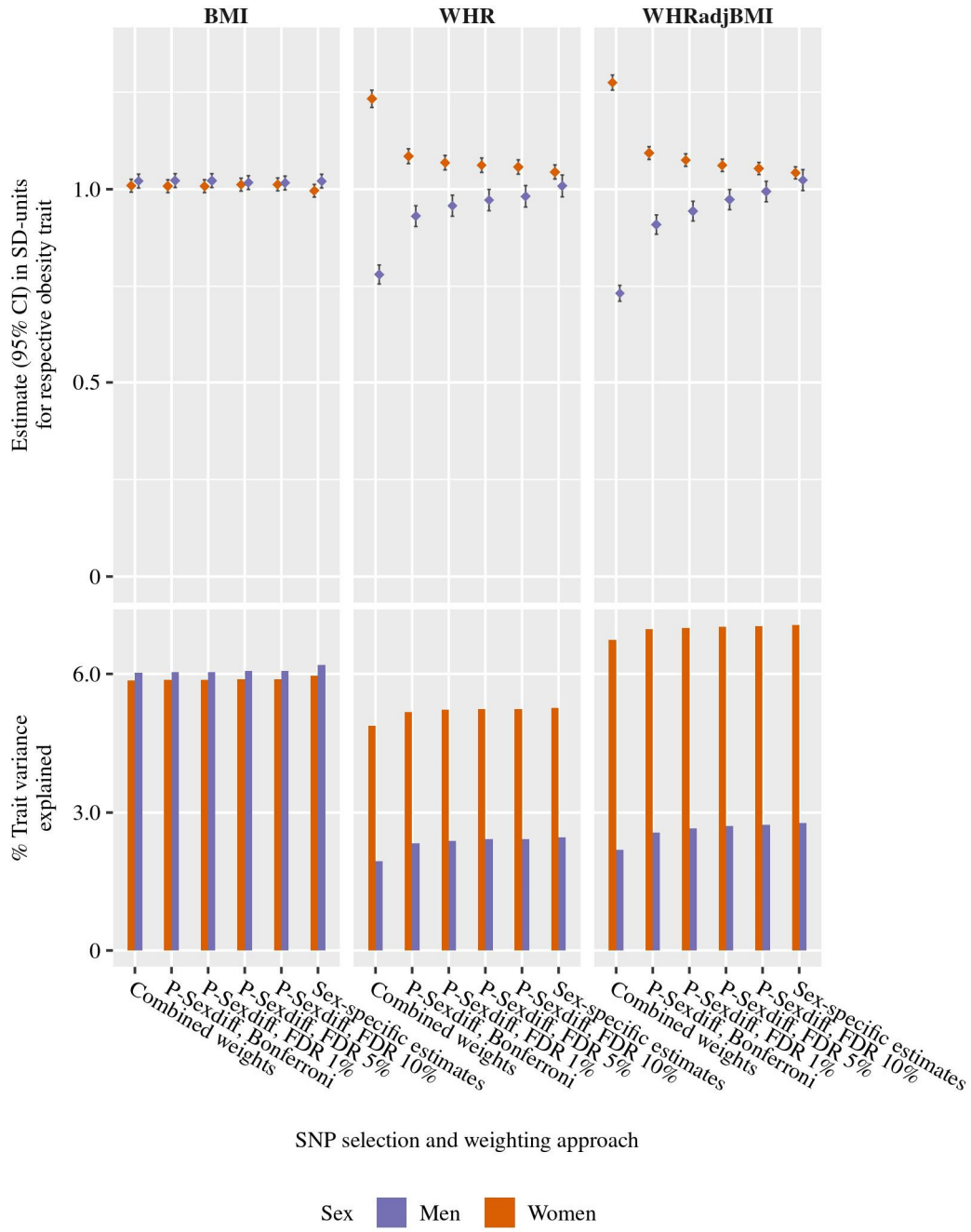


Figure 2.5. Trait variance explained and effect estimates with 95% confidence intervals (CIs) per 1-unit higher sex-specific GRS. Six of the approaches used to construct sex-specific GRSs contained the same SNPs, but with the SNPs weighted using different proportions of sex-combined versus sex-specific effect estimates. For the Combined weights instruments, sex-combined weights were used for all SNPs in both the men- and women-stratified analyses. For the Sex-specific estimates approach, all SNPs were weighted by their sex-specific effect estimates. For the P-Sexdiff approaches, the SNPs were weighted by their sex-specific estimate if the $P_{sexdiff}$-value for that SNP was below the specified threshold (Bonferroni, 1% FDR, 5% FDR, and 10% FDR, respectively), and otherwise weighted by its sex-combined estimate (Bonferroni, 1936; Benjamini and Hochberg, 1995). Thereafter, I performed sex-stratified association testing between the sex-specific GRSs and their respective obesity trait and computed the trait variance explained (Osborne, 2017c). Figure adapted (changed text and updated layout) from Censin, Peters, et al. (2019) and the corresponding pre-print (Censin, Bovijn, et al., 2019).

2.3.3. Sex-specific effects of obesity

2.3.3.1. Primary association testing of genetic risk scores for obesity with disease outcomes

The majority of the investigated disease outcomes were associated (P<0.001, for 0.05/51 investigated diseases overall) with at least one obesity GRS in the primary association testing using regression analysis, including CAD, CLD, COPD, NAFLD, stroke (overall and ischaemic, but not haemorrhagic), diabetes, renal failure, and lung cancer (Appendix Table 2.3) (Bonferroni, 1936). No obesity GRS were associated with dementia, haemorrhagic stroke, colorectal cancer, breast cancer, or infertility in any sex-strata.

2.3.3.2. Obesity is implicated as a risk factor in the majority of the leading causes of death

Obesity GRS to disease regression analyses that surpassed the multiple testing threshold in the primary association testing were taken forward for formal computation of MR estimates (Appendix Table 2.3). This included analyses involving CAD, CLD, COPD, lung cancer, NAFLD, renal failure, overall and ischaemic stroke, as well as both T1D and T2D. If a regression analysis was significant in only one sex, MR analysis was still performed in both sexes so that sexual heterogeneity could be assessed. The same multiple testing threshold (P<0.001, for 0.05/51 investigated trait-disease combinations overall) as in the regression analyses was used for the MR analyses (Bonferroni, 1936).

General obesity, as measured by BMI, was implicated as a risk factor for all the diseases brought forward for MR analysis in at least one sex-strata (Fig 2.6). Although the effect directions were consistent between men and women for all the BMI to disease MR analyses, not all combinations surpassed the multiple testing threshold in both sexes, including CLD, T1D, and lung cancer in men and ischaemic stroke in women (Bonferroni, 1936).

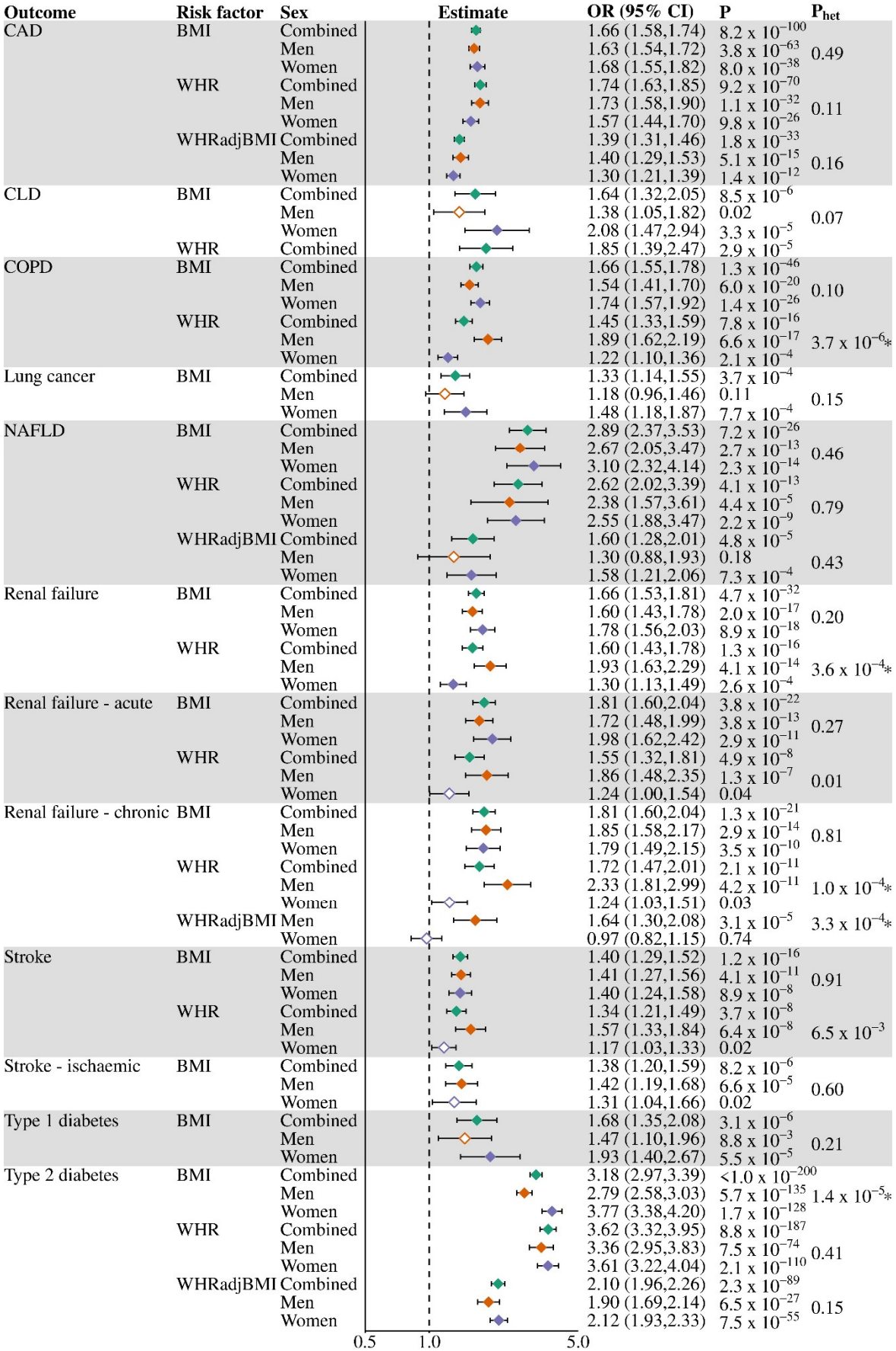


Figure 2.6. Sex-specific effects of obesity on leading causes of death. MR estimates computed using the ratio method and with effect estimates per 1 sex-specific SD higher obesity trait (Wald,

1940; Burgess and Thompson, 2015). Obesity trait-disease associations surpassing the P-value threshold (<0.001, for 0.05/51 obesity trait-disease combinations overall) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.001, for 0.05/48 obesity trait-disease combinations where sexual heterogeneity in the effect estimates were compared; no test for sexual heterogeneity performed for breast cancer) are marked with *. Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Figure adapted (updated layout and spelling) from Censin, Peters, et al. (2019) and the corresponding pre-print (Censin, Bovijn, et al., 2019).

Only a few diseases had robust support for abdominal obesity being implicated in the disease aetiology, with support of an effect in both the WHR and WHRadjBMI analyses. For CAD and T2D, there was evidence of abdominal obesity being a risk factor for the diseases in both men and women. The results also supported an effect of abdominal obesity on risk of NAFLD in both sexes, although the WHRadjBMI to NAFLD analysis did not surpass the multiple testing threshold in men. However, there was only robust evidence for abdominal obesity having an effect on chronic renal failure in men. In addition, there was some evidence for abdominal obesity (measured as WHR) being implicated in CLD, COPD, acute and overall renal failure, and overall stroke in at least one sex-strata.

No obesity trait seemed to have opposite effect directions on risk of disease in men versus women, that is, was protective against disease in one sex, while harmful in the other sex (Fig 2.6). However, the magnitude of the effect estimates differed (P_{het}<0.001, for 0.05/48 obesity trait-disease combinations evaluated for sexual heterogeneity) between men and women for several obesity trait-disease combinations (Bonferroni, 1936). Waist-related obesity traits had a higher effect estimate for risks of COPD (WHR $P_{het}=3.7\times10^{-6}$) and renal failure (WHR $P_{het}=3.6\times10^{-4}$), particularly chronic renal failure (WHR $P_{het}=1.0\times10^{-4}$ and WHRadjBMI $P_{het}=3.3\times10^{-4}$), in men than in women, whereas BMI seemed to increase risk of T2D more in women than in men (BMI $P_{het}=1.4\times10^{-5}$).

2.3.3.3. Obesity has strong effects on cardiometabolic risk factors

Obesity traits were associated (P<0.003, for 0.05/15 obesity trait-cardiometabolic risk factor combinations investigated in total) with increased risk of all cardiometabolic risk factors in both

men and women, although not all obesity traits were implicated for each cardiometabolic trait (Fig 2.7 and 2.8) (Bonferroni, 1936). For most obesity trait-risk factor combinations, the effect estimates were similar between men and women (multiple testing threshold set to P_{het}<0.003, for 0.05/15 obesity trait-cardiometabolic risk factor combinations evaluated for sexual heterogeneity) (Bonferroni, 1936). However, there was evidence of BMI increasing DBP more in women than in men ($P_{het}=5.2\times10^{-5}$), whereas BMI and WHR seemed to increase the risk of having been/being a smoker more in men than in women (BMI $P_{het}=8.4\times10^{-4}$ and WHR $P_{het}=2.2\times10^{-14}$).

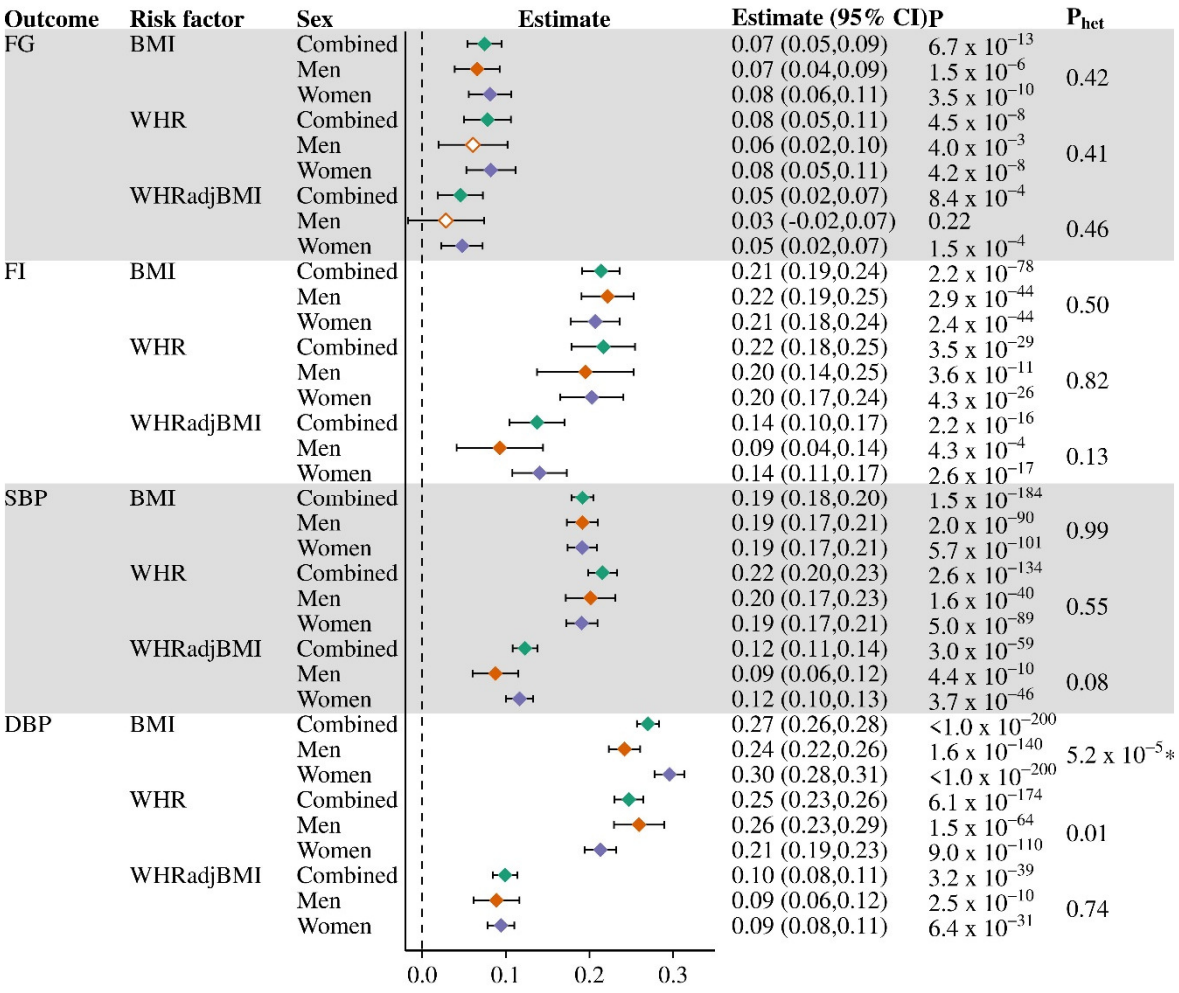


Figure 2.7. Sex-specific effects of obesity traits on cardiometabolic risk factors. MR estimates computed using the ratio method and with effect estimates per 1 sex-specific SD higher obesity trait (Wald, 1940; Burgess and Thompson, 2015). Obesity trait-cardiometabolic risk factor combinations surpassing the P-value threshold (<0.003, for 0.05/15 obesity trait-risk factor combinations investigated) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.003, for 0.05/15 obesity trait-risk factor combinations where sexual heterogeneity in the effect estimates were compared) are marked with * (Bonferroni, 1936). Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Figure adapted (updated layout) from Censin, Peters, et al. (2019) and the corresponding pre-print (Censin, Bovijn, et al., 2019).

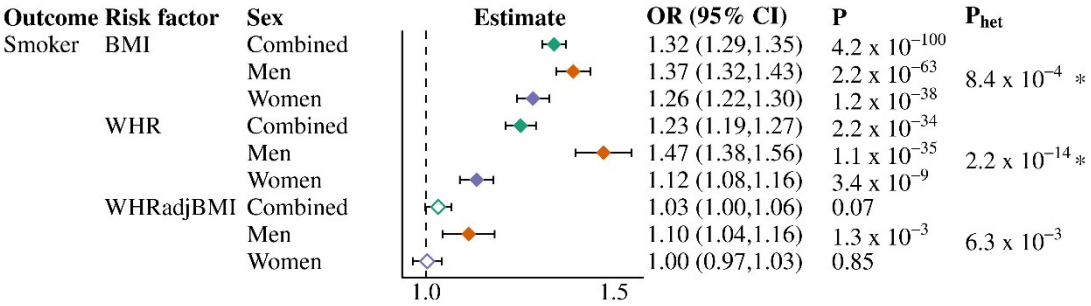


Figure 2.8. Sex-specific effects of obesity traits on smoking status. MR estimates computed using the ratio method and with effect estimates per 1 sex-specific SD higher obesity trait (Wald, 1940; Burgess and Thompson, 2015). Obesity trait-smoking status associations surpassing the P-value threshold (<0.003, for 0.05/15 obesity trait-risk factor combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.003, for 0.05/15 obesity trait-risk factor combinations where sexual heterogeneity in the effect estimates were compared) are marked with * (Bonferroni, 1936). Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Figure adapted (updated layout) from Censin, Peters, et al. (2019) and the corresponding pre-print (Censin, Bovijn, et al., 2019).

2.3.4. Sensitivity analyses

2.3.4.1. Mendelian randomisation analyses using additional weighting strategies

The externally weighted approach consisted of much fewer SNPs compared to the main analysis, and the unweighted GRSs were not weighted by their effect estimate. Still, all obesity trait-disease combinations that were significant (P<0.001) in the main analysis were at least nominally significant (P<0.05) using these weighting strategies (Appendix Figs 2.1-2.2).

The obesity trait-disease combinations that had evidence of sexual heterogeneity (P_{het}<0.001) in the main analysis also had at least nominal evidence (P_{het}<0.05) of sexual heterogeneity using these other weighting approaches, with one exception. There was no evidence of sexual heterogeneity in the effect of WHR on *overall* renal failure using the externally weighted approach (P_{het}=0.18). Still, the point estimate for men was higher than that for women regardless of weighting approach, and the P_{het}-values for a sex difference in the effect estimates of WHR on *chronic* renal failure were at least nominally significant (P_{het}<0.05) in all analyses.

For the obesity trait-disease combinations that had evidence of sexually heterogenous effect estimates, I also reconducted the analyses using different weighting approaches based on the sex-combined primary SNPs (approach 10, 11, and 12). All obesity trait-disease combinations and tests for sexual heterogeneity in the male-female effect estimates had at least nominal evidence ($P_{het}<0.05$) in support of sexual heterogeneity in these sensitivity analyses (Appendix Figs 2.3-2.5).

For the risk factor MRs, the obesity traits showed robust associations with the blood pressure traits using both externally weighted and unweighted GRSs, and with a robust sex difference in the BMI-DBP relationship ($P_{het}=2.1\times10^{-4}$ and $P_{het}=5.6\times10^{-4}$, respectively; Appendix Table 2.4). The obesity traits to smoking status MRs were largely supportive of an effect of the obesity traits on smoking propensity, but without conclusive evidence from all GRS approaches (Appendix Table 2.5).

2.3.4.2. Assessing the impact of normalisation approach and unit on sexual heterogeneity

The disease MR results were highly similar when computing MR estimates per 1 SD higher obesity trait computed jointly in men and women (Fig 2.9) and per 1 clinical unit higher obesity trait (kg/m^2 for BMI and per 0.01 higher ratio for WHR; Fig 2.10). The same obesity trait-disease combinations surpassed the multiple testing threshold ($P<0.001$) in these sensitivity analyses (Bonferroni, 1936). The sex differences in the effect of WHR on COPD, renal failure, and chronic renal failure, and WHRadjBMI on chronic renal failure, were robust (Fig 2.11). However, women did no longer have a significantly higher risk of T2D than men using the Sex-combined normalisation approach ($P_{het}=0.1$) nor when the computing effect estimates in clinical units ($P_{het}=0.33$).

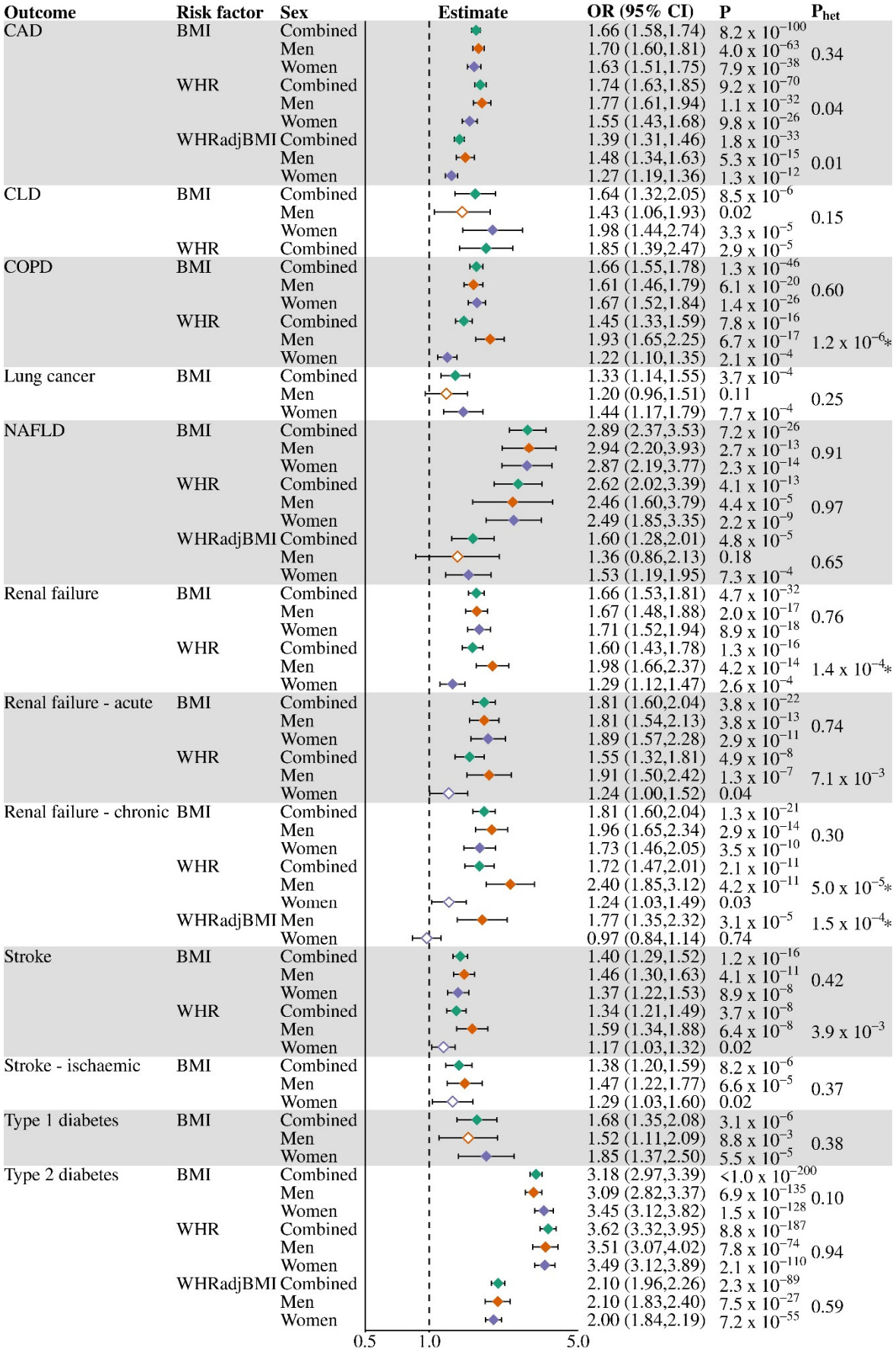


Figure 2.9. Sex-specific effects of obesity on leading causes of death using a sex-combined normalisation approach. MR estimates computed using the ratio method and with effect

estimates per 1 sex-combined SD higher obesity trait (Wald, 1940; Burgess and Thompson, 2015). Obesity trait-disease associations surpassing the P-value threshold (<0.001, for 0.05/51 obesity trait-disease combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.001, for 0.05/48 obesity trait-disease combinations where sexual heterogeneity in the effect estimates were compared; no test for sexual heterogeneity performed for breast cancer) are marked with * (Bonferroni, 1936). Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Layout copied from figures in Censin, Peters, et al. (2019) and the corresponding pre-print (Censin, Bovijn, et al., 2019).

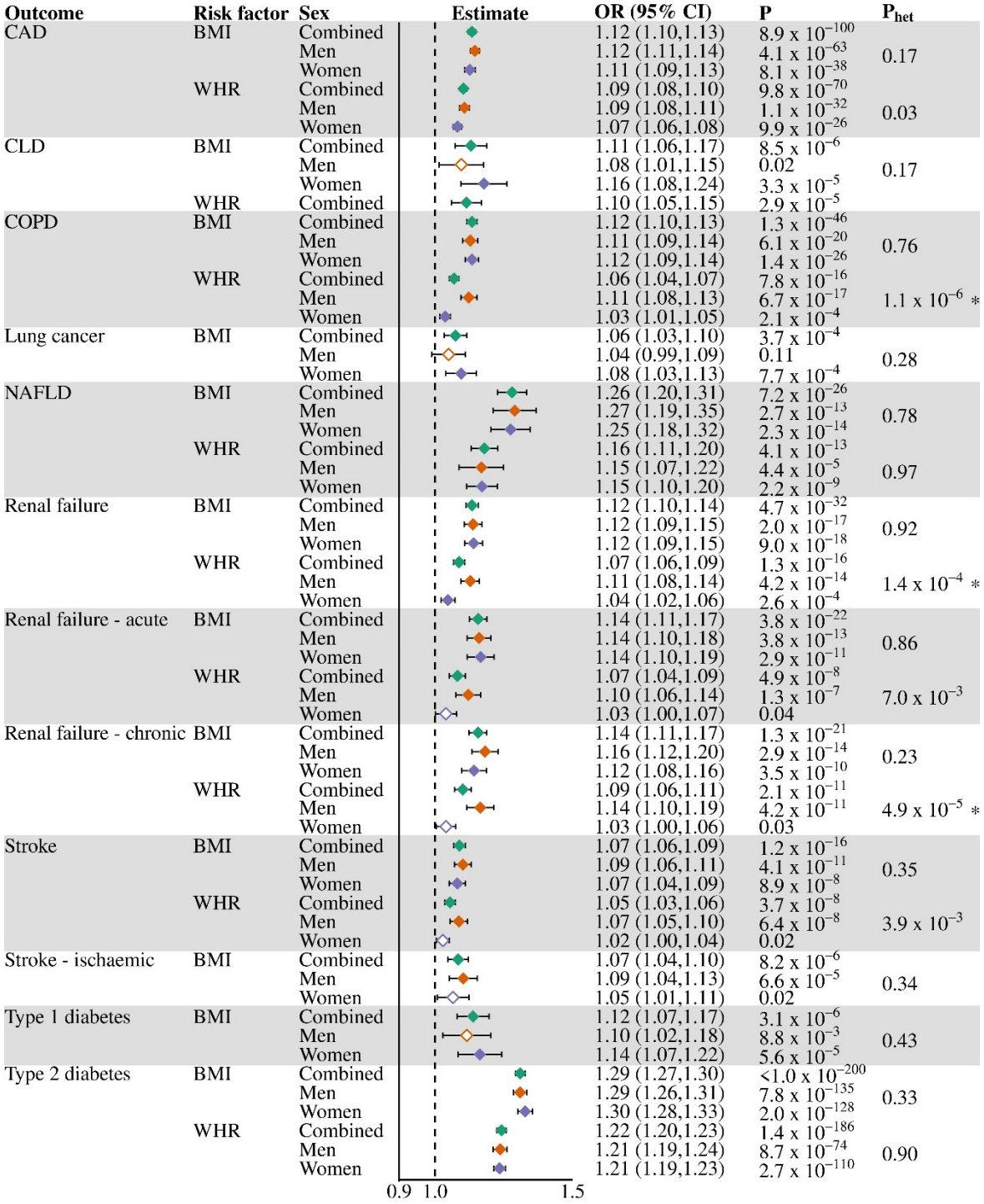


Figure 2.10. Sex-specific effects of obesity on leading causes of death for BMI and WHR in clinical units. MR estimates computed using the ratio method and with effect estimates per 1

kg/m2 higher BMI or 0.01 ratio higher WHR (Wald, 1940; Burgess and Thompson, 2015). Effect estimates per 1 additional kg/m^2 for BMI and per 0.01 higher ratio for WHR. Obesity trait-disease associations surpassing the P-value threshold (<0.001, for 0.05/51 obesity trait-disease combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.001, for 0.05/48 obesity trait-disease combinations where sexual heterogeneity in the effect estimates were compared; no test for sexual heterogeneity performed for breast cancer) are marked with * (Bonferroni, 1936). Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Layout copied from figures in Censin, Peters, et al. (2019) and the corresponding pre-print (Censin, Bovijn, et al., 2019).

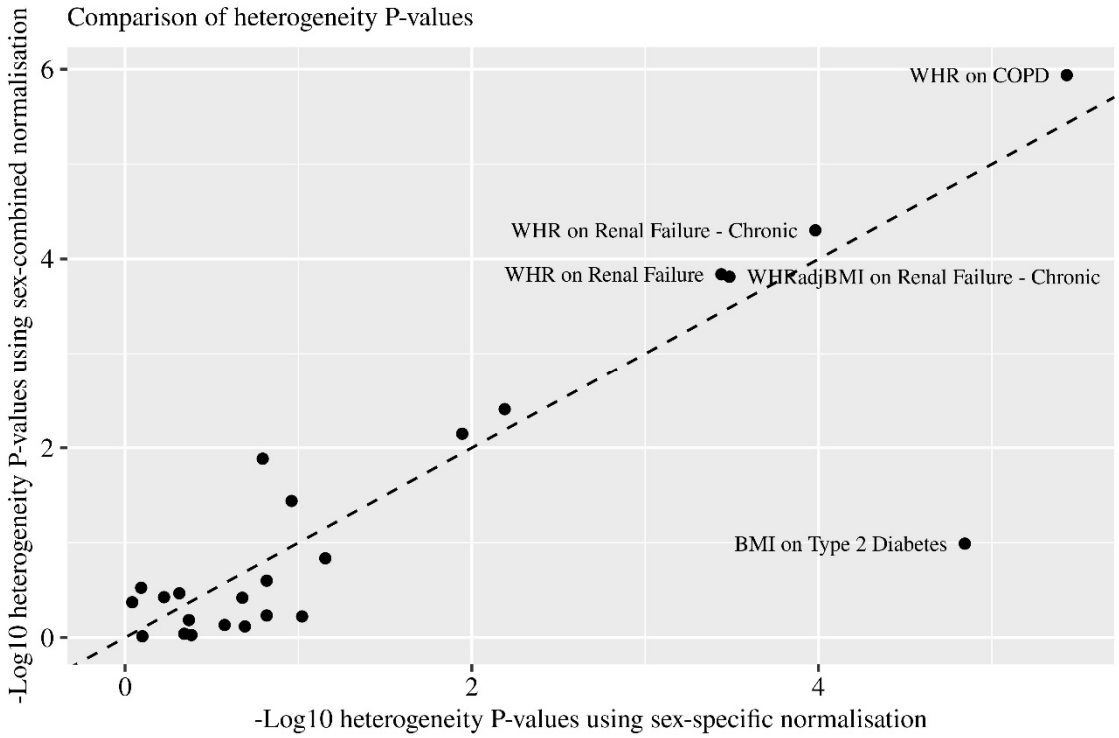


Figure 2.11. Comparison of -log10 P_{het}-values computed using a sex-specific normalisation approach and a sex-combined normalisation approach. P_{het}-values from MR analyses computed using the ratio method (Wald, 1940; Burgess and Thompson, 2015). P_{het}-values that surpassed the P_{het}-threshold (<0.001, for 0.05/48 obesity trait-disease combinations where sexual heterogeneity in the effect estimates were compared; no test for sexual heterogeneity performed for breast cancer) using the main sex-specific normalisation approach are labelled (Bonferroni, 1936).

There was still evidence in support of an effect of the obesity traits on SBP and DBP using a sex-combined normalisation approach (Fig 2.12), but there was no longer evidence of BMI increasing DBP to a larger extent in women than in men (P_{het}=0.64). However, there was evidence of WHR increasing DBP more in men than in women in these sensitivity analyses (P_{het}=3.6×10^{-4}). The effect estimates and sex differences in the effect of obesity traits on smoking status were robust to normalisation approach (Fig 2.13).

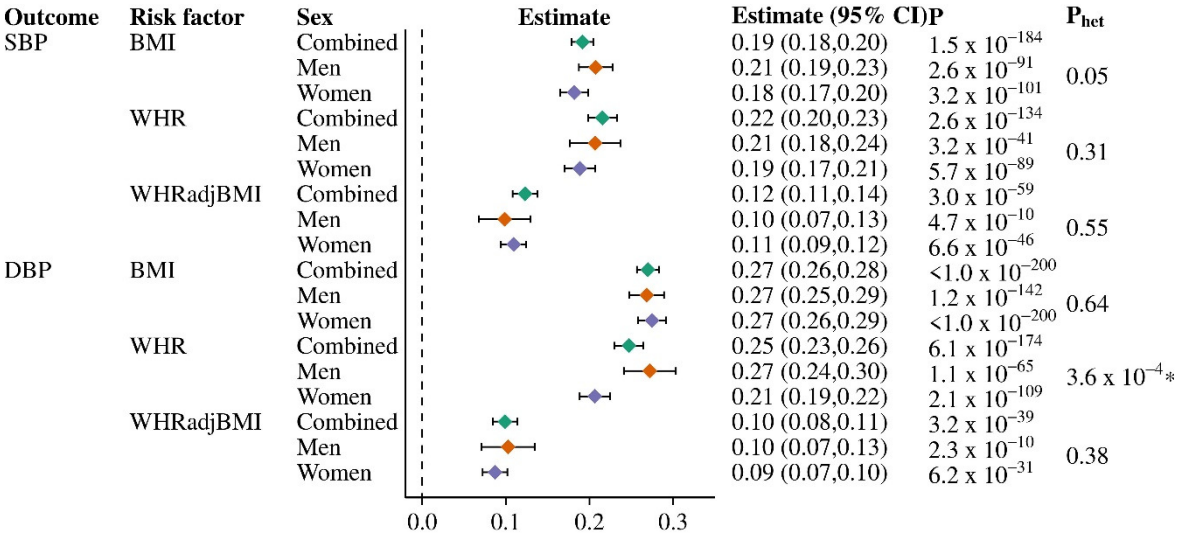


Figure 2.12. Sex-specific effects of obesity traits on SBP and DBP using a sex-combined normalisation approach. MR estimates computed using the ratio method and with effect estimates per 1 sex-combined SD higher obesity trait (Wald, 1940; Burgess and Thompson, 2015). Obesity trait-risk factor associations surpassing the P-value threshold (<0.003, for 0.05/15 obesity trait-risk factor combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.003, for 0.05/15 obesity trait-risk factor combinations where sexual heterogeneity in the effect estimates were compared) are marked with * (Bonferroni, 1936). Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Layout copied from figures in Censin, Peters, et al. (2019) and the corresponding pre-print (Censin, Bovijn, et al., 2019).

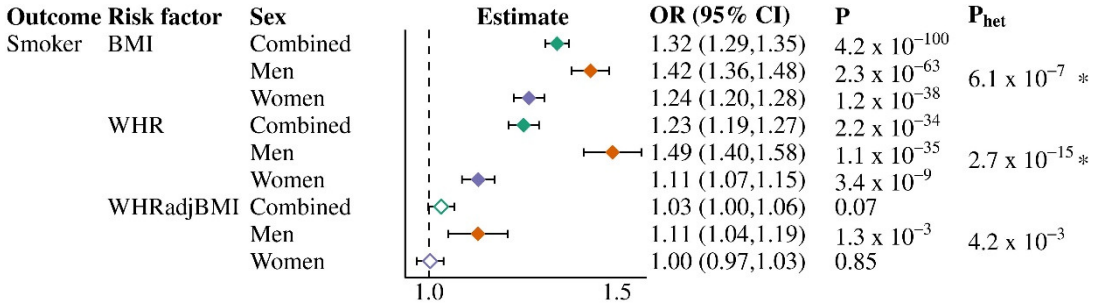


Figure 2.13. Sex-specific effects of obesity traits on smoking status using a sex-combined normalisation approach. MR estimates computed using the ratio method and with effect estimates per 1 sex-specific SD higher obesity trait (Wald, 1940; Burgess and Thompson, 2015). Obesity trait-smoking status associations surpassing the P-value threshold (<0.003, for 0.05/15 obesity trait-risk factor combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.003, for 0.05/15 obesity trait-risk factor combinations where sexual heterogeneity in the effect estimates were compared) are marked with * (Bonferroni, 1936). Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Layout copied from figures in Censin, Peters, et al. (2019) and the corresponding pre-print (Censin, Bovijn, et al., 2019).

2.3.4.3. Analyses performed in British ancestry subset

To reduce the risk of my results being due to population stratification (Berg *et al.*, 2019; Haworth *et al.*, 2019), I repeated the analyses in the British ancestry subset. These results strongly supported

the main findings (Appendix Fig. 2.6). Even though the sample size was 10% lower in the British ancestry subset, all obesity trait-disease outcomes and sex differences remained associated ($P<0.001$) in the sensitivity analysis with three exceptions (Bonferroni, 1936). WHR was no longer significantly associated with increased COPD and renal failure in women and increased ischaemic stroke in men, although the P-values remained nominally ($P<0.05$) significant in the British ancestry subset (WHR-COPD in women: $P=0.002$; WHR-renal failure in women: $P=0.001$; WHR-ischaemic stroke in men: $P=0.001$).

Similarly, all obesity traits remained significantly ($P<0.003$) associated with the blood pressure traits in the White British ancestry subsample, and there was still evidence of sexual heterogeneity in the effect of BMI on DBP (Appendix Table 2.6). The obesity trait to smoking status MRs were also at least nominally supported ($P<0.05$ and $P_{het}<0.05$) by the sensitivity analyses using the White British ancestry subset and the main GRSs approach (Appendix Table 2.5).

2.3.4.4. Analyses performed using the same number of cases and controls

For some diseases, the number of cases differed dramatically between men and women (Table 2.2). To minimise the risk of sex-differential power affecting my results, I conducted logistic regressions with the same number of both cases and controls in the two sexes, which yielded highly similar results to the main analyses (Appendix Table 2.7). All the obesity trait-disease combinations with evidence of sexual heterogeneity in the MR analyses also had at least nominal ($P_{het}<0.05$) evidence of sexual heterogeneity in this sensitivity analysis.

2.3.4.5. Analyses performed using stricter diabetes case criteria

The diabetes case algorithm classified individuals with diabetes as either "probable" or "possible" to have T1D and T2D (Eastwood *et al.*, 2016). To minimise the risk of misclassification of diabetes type affecting my results, I reconducted the MR analyses assessing the effect of obesity on T1D and T2D using only "probable" cases as a sensitivity analysis. Results using "probable" T2D cases strongly supported ($P<0.001$) the evidence of all obesity traits increasing risk of T2D in all sex-strata, as well as BMI increasing the risk of T2D more in women than in men ($P_{het}=3.5\times10^{-5}$;

Fig 2.14) (Bonferroni, 1936). Use of stricter diabetes criteria also yielded similar estimates for the effect of BMI on T1D risk across the sex-strata. There was strong evidence of BMI increasing the risk of T1D in women and sex-combined, but only nominal evidence (P=0.02) in men, similar to the main analysis.

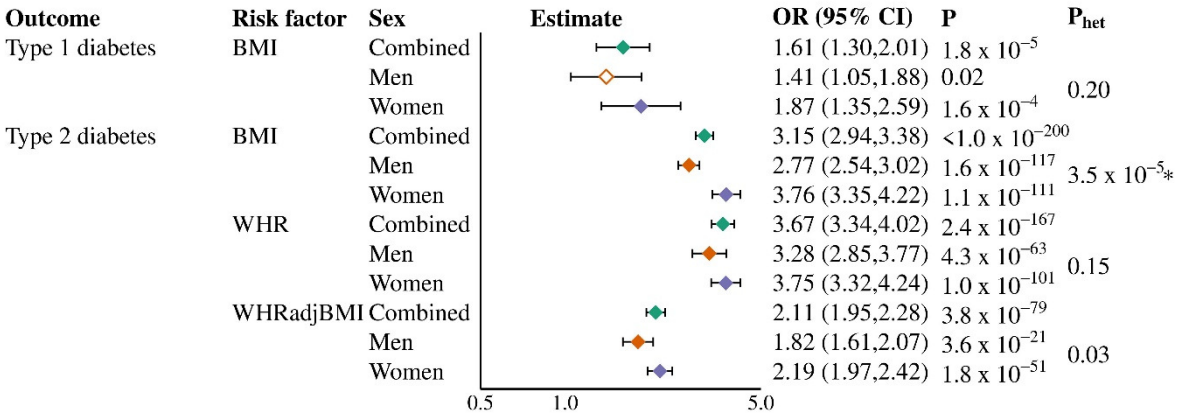


Figure 2.14. Sex-specific effects of obesity traits on diabetes risk using stricter diabetes case criteria. MR estimates computed using the ratio method and with effect estimates per 1 sex-specific SD higher obesity trait (Wald, 1940; Burgess and Thompson, 2015). Obesity trait-smoking status associations surpassing the P-value threshold (<0.003, for 0.05/15 obesity trait-risk factor combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.003, for 0.05/15 obesity trait-risk factor combinations where sexual heterogeneity in the effect estimates were compared) are marked with * (Bonferroni, 1936). Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Layout copied from figures in Censin, Peters, et al. (2019) and the corresponding pre-print (Censin, Bovijn, et al., 2019).

2.3.4.6. Pleiotropy-robust sensitivity analyses

I also reperformed the disease MRs using more pleiotropy-robust methods for the main sex-specific estimates and externally weighted approaches, to decrease the risk of pleiotropic instruments biasing my results (Swanson and Hernán, 2013; Burgess and Thompson, 2015; Burgess, Davey Smith, *et al.*, 2020). Effect estimates computed using the median-weighted method were generally less precise than those computed using the ratio method in the main analysis (Fig 2.15 and Appendix Fig 2.7) (Wald, 1940; Burgess and Thompson, 2015; Bowden *et al.*, 2016). While most obesity trait-disease combinations remained nominally significant (P<0.05) in these sensitivity analyses, the estimate of the effect of WHR on COPD in women was no longer directionally consistent using the main Sex-specific estimates approach.

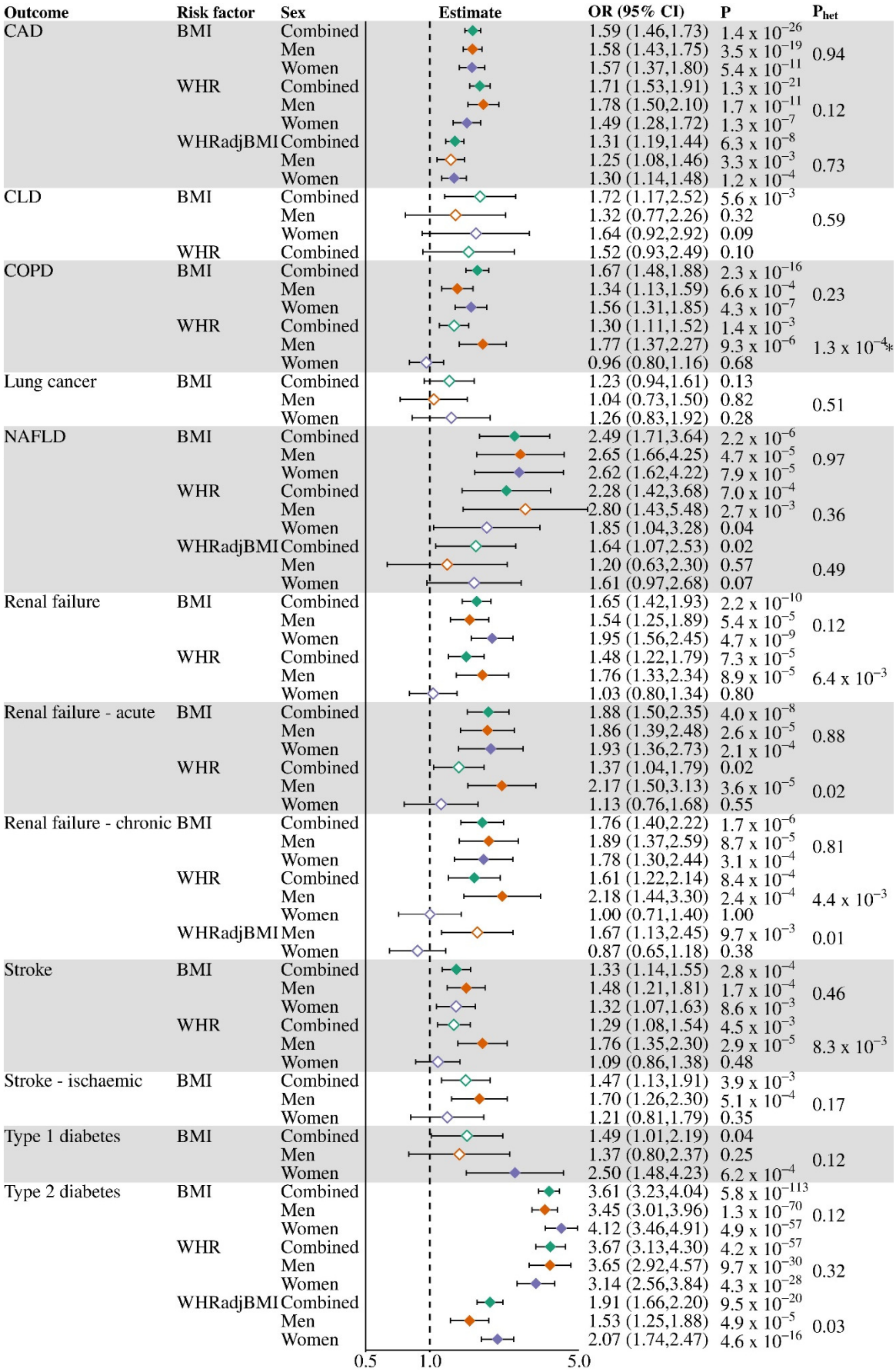


Figure 2.15. Sex-specific effects of obesity on leading causes of death, using the weighted median method. MR estimates computed using the weighted median method and with effect

estimates per 1 sex-specific SD higher obesity trait (Burgess and Thompson, 2015; Bowden *et al.*, 2016). Obesity trait-disease associations surpassing the P-value threshold (<0.001, for 0.05/51 obesity trait-disease combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.001, for 0.05/48 obesity trait-disease combinations where sexual heterogeneity in the effect estimates were compared; no test for sexual heterogeneity performed for breast cancer) are marked with * (Bonferroni, 1936). Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Layout copied from figures in Censin, Peters, et al. (2019) and the corresponding pre-print (Censin, Bovijn, et al., 2019).

Analyses using the MR-Egger method yielded much less precise effect estimates (Appendix Table 2.8 and Appendix Fig 2.8-2.9) (Bowden, Davey Smith and Burgess, 2015). In three cases, the effect estimates were not directionally consistent, including for WHR on COPD in women, as well as BMI on lung cancer and WHR on acute renal failure in the sex-combined analyses. For a few obesity trait-disease combinations, the MR-Egger intercept test gave evidence for directional pleiotropy (P<0.002, for 0.05/25 obesity trait-outcome combinations), with a lower magnitude of the MR-Egger point estimate in these analyses (including for WHR-CAD in sex-combined, BMI-COPD in men, WHR-COPD in women, WHR-renal failure in men, and WHR-T2D in women) (Bonferroni, 1936; Bowden, Davey Smith and Burgess, 2015).

Analyses using the externally weighted SNPs and the more pleiotropy-robust MR-methods rendered some support for WHR increasing risk of COPD more in men than in women (MR-Egger method: $P_{het}=5.5\times10^{-3}$; Weighted median method: $P_{het}=0.08$) (Bowden, Davey Smith and Burgess, 2015; Bowden *et al.*, 2016). There was less evidence that the obesity traits increased renal failure more in men than in women using the pleiotropy-robust methods, although the point estimates were consistently higher for men for the effect of WHR on renal failure (weighted median $P_{het}=0.37$ and MR-Egger $P_{het}=0.52$) and chronic renal failure (weighted median $P_{het}=0.05$ and MR-Egger $P_{het}=0.25$), and WHRadjBMI on chronic renal failure (weighted median $P_{het}=0.13$ and MR-Egger $P_{het}=0.81$) (Bowden, Davey Smith and Burgess, 2015; Bowden *et al.*, 2016). These sensitivity analyses using pleiotropy robust MR methods and externally weighted GRSs were more inconclusive regarding a sex difference in the male versus female effect estimates of the effect of BMI on risk of T2D. The point estimates were almost identical between men and women using the weighted median method (men OR 3.33, 95% CI 2.84-3.90, and women OR 3.29, 95% CI 2.72-

3.99, P_{het}=0.93), whereas the sex-specific MR-Egger estimates gave some support of BMI increasing T2D risk more in women than in men (men OR 1.53, 95% CI 1.14-2.05, and women OR 2.44, 95% CI 1.69-3.51, P_{het}=0.05) (Bowden, Davey Smith and Burgess, 2015; Bowden *et al.*, 2016).

The MR-Egger and weighted median results for the effect of obesity on FG and FI supported the main findings (Appendix Table 2.9) (Bowden, Davey Smith and Burgess, 2015; Bowden *et al.*, 2016). All causal effects identified in the main analysis had consistent effect directions in these sensitivity analyses, and with no clear evidence of directional pleiotropy in the MR-Egger intercept test.

2.4. Discussion and Conclusion

My first thesis aim was to determine the effect of obesity on leading causes of death in men versus women. In this chapter, my work has implicated obesity in the majority of leading mortality causes in both men and women, but also demonstrated that obesity traits may confer different magnitudes of risk for COPD, renal failure and T2D in men versus women. These findings have important implications for epidemiologic research into the consequences of obesity, and suggest that sex-specific preventative measures may be warranted.

2.4.1. Instruments

My first research objective for this chapter was to evaluate different approaches to construct sex-specific GRSs for obesity. My results indicate that care needs to be taken when constructing sex-specific GRSs for certain traits. Prior to conducting the analyses, my hypothesis was that using sex-combined weights for SNPs without evidence of any sex difference in the effect estimates would improve the GRSs because of the larger sex-combined sample size. However, my data indicate that use of sex-combined weights could cause GRSs to systematically under/overestimate the effect of the GRSs on their respective obesity trait in each sex. In the original GWAS on

obesity, Pulit *et al.* (2019) found that most sex-differential SNPs for abdominal obesity had higher effect estimates in women. It is possible that there is a similar pattern of SNPs having larger effects in women even if they do not surpass the $P_{sexdiff}$-threshold. If so, the use of sex-combined weights would lead to SNPs being underestimated in women and overestimated in men, which could explain my results. In addition, my results demonstrated that weighting all SNPs by their sex-specific effect estimates rather than sex-combined weights could improve instrument strength and results in higher trait variances explained and F-statistics (Burgess and Thompson, 2011; Osborne, 2017c).

To decrease the risk of bias from weak instruments, I combined the obesity-associated SNPs into single sex-specific GRSs for each obesity trait. Combining SNPs into a single instrument will increase the F-statistic and thereby decrease the risk of weak instrument bias (Burgess and Thompson, 2013; Burgess, Dudbridge and Thompson, 2016). Commonly, an F-statistic <10 is used to define a weak instrument (Burgess and Thompson, 2015). The lowest F-statistic of any of the instruments used in the main analyses had an F-statistic of 4,921 and a trait variance explained of 2.47%. In summary, the use of a single instrument and the high F-statistics indicate that any bias from weak instruments should be negligible (Burgess, Davies and Thompson, 2016).

2.4.2. Diabetes and glycaemic traits

My results indicated that BMI may increase risk of T2D more in women than in men. Yet, this sex difference was only present when computing MR estimates per 1-unit higher sex-specific SD of BMI. There was no sex difference in effect estimates computed per 1 kg/m^2 or per 1-unit higher sex-combined SD of BMI. This discrepancy is likely explained by the fact that 1 SD of BMI corresponded to a higher amount of kg/m^2 in women than in men.

Previous observational studies have indicated that WHR may be a better predictor of risk of T2D in women than in men (Meisinger *et al.*, 2006; Wannamethee *et al.*, 2010). Yet, similar to my results,

two additional studies assessing how genetic predisposition to WHRadjBMI affect risk of T2D in men and women found no evidence of sexual heterogeneity (T. Huang *et al.*, 2015; Lotta *et al.*, 2018). Observational studies have also indicated that higher BMI may be associated with lower insulin sensitivity in men than in women (Sierra-Johnson *et al.*, 2004; Masharani, Goldfine and Youngren, 2009). However, I found no evidence of a sex difference in the effect of any of the obesity traits on FG and FI.

BMI has previously been implicated in risk of T1D (Verbeeten *et al.*, 2011), and a previous MR study by colleagues and I found evidence for childhood BMI increasing risk of T1D (Censin *et al.*, 2017). My results in this chapter replicated these previous findings. However, some genetic variants associated with adult BMI are also associated with childhood BMI (Felix *et al.*, 2016; Censin *et al.*, 2017), and it is possible that my GRSs for adult obesity in this chapter are capturing effects from childhood obesity rather than adulthood obesity. Another potential issue in the findings concerning T1D and T2D is misclassification of diabetes type among the cases, however my sensitivity analyses using stricter diabetes case definitions strongly supported the main findings.

2.4.3. Cardiovascular disease

My results suggested a relationship between BMI and higher risk of overall stroke in both men and women, as well as for ischaemic stroke in men. In addition, there was evidence of WHR increasing risk of overall stroke in men. However, there was only nominal evidence of an effect of BMI on ischaemic stroke and WHR on overall stroke in women. The lower strength of association in women could be a power issue, as there were only about half as many female ischaemic stroke cases as male cases. Yet, sensitivity analyses performed using the same number of cases and controls in both sexes yielded similar results, and there was nominal evidence of sexual heterogeneity in the WHR to stroke effect estimates. Previous observational studies have indicated that BMI is associated with increased risk of stroke, but the results have been inconclusive as to

whether WHR is associated with risk of stroke in men and women (Hu *et al.*, 2007; The Global Burden of Metabolic Risk Factors for Chronic Diseases Collaboration, 2014; Abete *et al.*, 2015). My results suggest that WHR may increase risk of stroke in both sexes, but that WHR may increase risk of overall stroke to a larger degree in men – but more studies are needed as the association statistics did not surpass the multiple testing threshold (Bonferroni, 1936).

All measured obesity traits increased risk of CAD to a similar degree in both men and women. The associations of the obesity traits on risk of CAD are consistent with previous observational and MR studies, although previous observational studies investigating sex differences in the effect of obesity traits on cardiovascular disease have been inconclusive (Dagenais *et al.*, 2005; Taylor *et al.*, 2010; Dale *et al.*, 2017; Emdin *et al.*, 2017; Lyall *et al.*, 2017; Lotta *et al.*, 2018; Rost *et al.*, 2018; Millard *et al.*, 2019). A previous MR study by Lotta *et al.* (2018) could not identify any sex difference in the effect of WHRadjBMI on risk of CAD either, similar to my findings.

2.4.4. Lung disease and smoking

I found evidence of BMI and WHR increasing the risk of COPD in both sexes, but that WHR may increase the risk of COPD to a higher degree in men than in women. Analyses using externally weighted SNPs and the more pleiotropy-robust MR-Egger method largely supported WHR increasing risk of COPD more in men than in women (Bowden, Davey Smith and Burgess, 2015). There was also evidence of BMI increasing the risk of lung cancer in women but not in men, but without evidence of sexual heterogeneity in the causal effect estimates, thus largely replicating a previous MR study (Carreras-Torres *et al.*, 2017).

A likely explanation of how the obesity traits affect risk of COPD and lung cancer is through smoking. I found that BMI and WHR increased smoking propensity in both sexes, but to a higher degree in men than in women. Previous studies have indicated that obesity may cause individuals to smoke more, possibly in an attempt to lose weight (Fulkerson and French, 2003; Chiolero *et al.*,

2008). Observational studies have indicated that smoking may lower body weight, but that higher BMI may also result in increased smoking (Rásky, Stronegger and Freidl, 1996; Audrain-McGovern and Benowitz, 2011; Morris *et al.*, 2015; Carreras-Torres *et al.*, 2017). It is thus possible that BMI affects COPD and lung cancer not only through physiological effects but also through changed smoking behaviour.

In spite of BMI previously being linked to risk of COPD in observational studies, obesity is not an established risk factor for the disease (Vozoris and O'Donnell, 2012; Vestbo *et al.*, 2013; Hanson *et al.*, 2014; Millard *et al.*, 2019). My work confirms obesity as a risk factor for both lung cancer and COPD. In addition, there has previously been a lack of research into sex-specific effects of obesity on smoking propensity, COPD, and lung cancer using an MR design. My results are thus novel in demonstrating that obesity traits may increase risks of COPD and smoking propensity more in men than in women using genetic epidemiology. However, it is possible that my results are affected by my smoking phenotype not capturing smoking to a similar degree in both sexes, and more studies are needed to confirm these results.

2.4.5. Renal failure

Previous MR studies investigating the effect of obesity on kidney diseases have been inconclusive, and have not investigated sex-specific effects (Todd *et al.*, 2015; Emdin *et al.*, 2017; Geng *et al.*, 2018; van Zuydam *et al.*, 2018; Millard *et al.*, 2019). My results indicate that obesity plays a role in both acute and chronic renal failure. In addition, I found evidence of abdominal obesity increasing risk of renal failure (especially chronic renal failure) to a larger extent in men than in women. However, the results were not unequivocally supported by sensitivity analyses and the reason for any sex difference is unclear. Obesity has been suggested to affect renal disease through several pathophysiological mechanisms, including higher blood pressure, increased risk of T2D, and through structural modifications in the kidney (Tsuboi *et al.*, 2012; Hall *et al.*, 2014; Kovesdy

et al., 2017; van Zuydam *et al.*, 2018). Yet, whether these mechanisms explain the observed sex differences in the effect of obesity on renal failure is not known.

2.4.6. Liver disease

In addition to analysing the effects of obesity on the leading causes of death, I also investigated the effect of the obesity traits on NAFLD and CLD. NAFLD and CLD are increasing causes of morbidity and mortality, and CLD can be caused by NAFLD (Vernon, Baranova and Younossi, 2011; Williams *et al.*, 2011; Chalasani *et al.*, 2012; Estes *et al.*, 2018). My results demonstrated obesity as a risk factor in NAFLD and CLD development, similar to a previous MR study that found BMI to increase liver fat content (Stender *et al.*, 2017). There was no evidence of a sex difference in the effect of the obesity traits on NAFLD and CLD, and my results thus emphasise obesity is an important risk factor for these diseases for both men and women.

2.4.7. Other diseases

There was no association between the GRSs for obesity and dementia, colorectal cancer, breast cancer in women (not investigated in men), haemorrhagic stroke, or infertility. Consequently, I did not take them forward for computation of formal MR estimates. There are several possible explanations to the null results, besides the obvious explanation that obesity is not causally related to these diseases.

Causal effects can be obscured for several reasons. For example, both the National Institute of Health in the US and the National Health Services in the UK lists obesity as a risk factor for breast cancer (National Health Service, 2019; National Cancer Institute at the National Institutes of Health, 2020). Yet, it is unclear if obesity increases the risks of both pre- and postmenopausal breast cancer. Previous studies generally agree that obesity is protective for pre-menopausal breast cancer, but have been inconclusive about whether obesity increases or decreases the risk of post-menopausal breast cancer (van den Brandt *et al.*, 2000; Guo *et al.*, 2016; Al-Ajmi *et al.*, 2018; Shu

et al., 2018). Thus, it could be that the effect of BMI on pre- versus post-menopausal breast cancer counteract each other in my analyses. Yet, two previous MR studies found a protective effect of BMI on *both* pre- and postmenopausal breast cancer risk, although these MR studies used a more limited set of SNPs than those used in my analyses (Guo *et al.*, 2016; Shu *et al.*, 2018). These discrepancies between previous studies warrant future studies to determine the exact effect of different obesity measures on breast cancer subtypes. For example, as suggested by Guo *et al.* (2016), effects could be dependent on weight and weight gain at certain time points in life.

For some of the disease outcomes, I may have had too few cases to detect a causal effect. For example, previous MR studies that have identified an effect of BMI on risk of colorectal cancer have had access to larger sample sizes (Thrift *et al.*, 2015; Jarvis *et al.*, 2016). I also had relatively few cases for dementia, infertility, and haemorrhagic stroke, wherefore these analyses in particular may have had insufficient statistical power to detect any causal effects.

2.4.8. Mendelian randomisation and sensitivity analyses

I reconducted the MR analyses using an externally weighted and unweighted approach (Burgess and Thompson, 2013; Burgess, Dudbridge and Thompson, 2016). I also reperformed the disease MRs with evidence of sexual heterogeneity using sex-combined primary SNPs, to limit the risk of my results being affected by one sex having more GWAS-significant index SNPs than the other sex (Pulit *et al.*, 2019). These sensitivity analyses rendered further support of the main conclusions.

I additionally assessed the effect of computing effect estimates per 1-unit higher sex-combined SD of the obesity trait and per clinical unit (kg/m^2 for BMI and ratio for WHR). While 1 SD corresponded to roughly equal amounts of WHR in men and women, 1 SD of BMI corresponded to more kg/m^2 in women than in men. I found that most sex-differential causal relationships between obesity traits and outcomes were robust to choice of unit, with the exception of BMI on T2D and DBP. In the article published with these results (Censin, Peters, *et al.*, 2019), we only included

results comparing men and women per 1 sex-specific SD higher obesity trait. This was later criticised in a pre-print by senior authors Nivard, Perry and Ganna, even though the authors conceded that *"... the differences in the variance of BMI between males and females likely reflects the distribution in the general population..."* (Pirastu *et al.*, 2020). Yet, in an MR study of testosterone with partly overlapping senior authors, they have also chosen to normalise testosterone sex-stratified and to contrast men and women MR estimates (Ruth *et al.*, 2020).

Obesity trait distributions differ between men and women in the UK population (Fry *et al.*, 2017). If the distributions of a trait differ between the sexes, computing effect estimates per 1 unit higher sex-combined SD of an obesity trait would not take any sex-differential trait distributions into account. In addition, for traits where the variance differs a lot between the sexes, a classical RINT of the residuals (Bliss, 1967) performed sex-combined could result in the sex with the larger variance having a seemingly binormal distribution. Thus, in my opinion, it seems fairer to compare men and women per 1-unit higher sex-standardised SD as a main analysis, and to further tease apart sex differences by supplementing with sensitivity analyses performed in other units.

However, Pirastu *et al.* (2020) raise another important issue; the potential of sex-differential participation bias, in essence a form of sex-differential collider bias. In brief, Pirastu *et al.* (2020) perform a GWAS of sex in up to 3,3 million men and women and identify several loci that seemingly associate with sex. The authors suggest that if, for example, obesity is sex-differentially associated with the likelihood of participating in a study, genetic variants that associate with obesity will seem associated with sex (Pirastu *et al.*, 2020). Indeed, genetic variants in the well-known obesity-associated locus *FTO* come up as associated with sex (Dina *et al.*, 2007; Frayling *et al.*, 2007; Pirastu *et al.*, 2020). Pirastu *et al.* (2020) also perform simulation studies using various settings, and suggest that sex-differential participation bias could cause the sexual heterogeneity in the BMI to T2D effect estimates that we observe (Censin, Peters, *et al.*, 2019). Yet, while it seems likely that sex-differential bias exists, it seems unlikely that it would cause such substantial bias that it would explain my results. In the simulations, Pirastu *et al.* (2020) vary two simulation

parameters in a scenario of no true sexual heterogeneity; OR and K. OR represents the degree of participation bias (the more OR deviates from 1, the higher the degree of participation bias) and K the degree of sex-differential bias (the lower/more negative values of K, the higher degree of sex-differential participation bias) (Pirastu *et al.*, 2020). The simulation parameters are described as being varied between 0.33 to 3 for OR and -0.5 to 1.5 for K (Pirastu *et al.*, 2020). For educational attainment, the authors estimate OR to 1.25 and K to 0.7 based on census data (Pirastu *et al.*, 2020). However, Pirastu *et al.* (2020) are only able to detect sex-differential bias causing sex differences surpassing my multiple testing P_{het}-threshold ($P_{het}<0.001$) in the most extreme simulation scenarios (Bonferroni, 1936). They highlight the results for K=-1 (more extreme than the reported simulation range described in their methods), and even then it is only in the simulation with the most extreme value of OR (0.33) that the P_{het}-value surpasses my multiple testing threshold (Bonferroni, 1936). In almost all other simulation scenarios, the P_{het}-value is not even nominally significant (Pirastu *et al.*, 2020). Using similar settings as those suggested for education, P_{het} is close to 1 (Pirastu *et al.*, 2020). In addition, another article by Munafó *et al.* (2018) also perform simulations to assess how different degrees of participation bias would cause deviations from mean trait levels in the underlying populations. Simulations are then performed based on different degrees of participation bias and trait variance explained (Munafò *et al.*, 2018). If the true population men and women SD are close to the ones I computed in the UKBB, there is a very slight deviation of mean UKBB BMI values compared to UK census data (absolute difference 0.6 kg/m^2 in women and 0.7 kg/m^2 in men, well below 0.2 SD for both sexes) (Fry *et al.*, 2017). In the simulations by Munafó *et al.* (2018), a trait difference between the true population and the study population of 0.2 SD corresponds to the weakest participation bias simulation statistic of OR=1.2. For simulation scenarios with this degree or participation bias, confidence intervals almost always contain the true value and the bias is very minor (Munafò *et al.*, 2018). Thus, even though sex-differential participation bias potentially could affect my results, it seems unlikely that such a bias would be substantial enough to explain the sexual heterogeneity in the causal estimates in my work.

There has also been concerns of both GIANT and the UKBB being affected by population stratification (Berg *et al.*, 2019; Haworth *et al.*, 2019). For example, Haworth *et al.* (2019) found that GRSs and PRSs of BMI using data from GIANT may associate with birth location in the UKBB. However, the degree of bias decreased when only using GWAS-significant SNPs in the GRSs (Haworth *et al.*, 2019). Haworth *et al.* (2019) found no association ($P<0.05$) with birth location for a weighted GRS using GWAS-significant SNPs and adjusting for array type, 10 PCs and assessment centre in the UKBB. I therefore adjusted for these variables in my analyses. In addition, I repeated the analyses – using the same adjustments – in the "White British" UKBB subset, with virtually the same results. Thus, whereas bias from population stratification is difficult to completely rule out, it seems unlikely that it substantially biases my results.

I was also concerned about the different numbers of men and women disease cases affecting my results, given that this would likely affect the power in the analyses. I therefore repeated the logistic regressions – essentially the numerator in the MR estimate – using the same number of men and women cases and controls. These results supported the main findings.

Another issue I considered was potential bias from misclassification of diabetes type. Even though type 1 and type 2 diabetes (and rarer diabetes types such as gestational diabetes) share common features, pathophysiological mechanisms are considered to be different (The World Health Organization, 2020a). I thus also performed the analyses restricting my type 1 and type 2 diabetes cases to only probable (as opposed to possible and probable) cases (Eastwood *et al.*, 2016). These sensitivity analyses using stricter diabetes case definitions yielded very similar results to those using more lenient definitions.

I also performed analyses (MR-Egger and weighted median methods) more robust to pleiotropy, using both the SNPs and weights included in the main instruments and externally weighted SNPs (Bowden, Davey Smith and Burgess, 2015; Bowden *et al.*, 2016). MR-Egger generally has much lower power than other MR-methods, and the MR-Egger estimates were imprecise compared to the

other methods (Bowden, Davey Smith and Burgess, 2015). It has also been noted that MR-Egger might considerably exacerbate weak instrument bias in one-sample or subsample MR, therefore the analyses using the main instruments and MR-Egger should be interpreted with great care (Burgess, Davies and Thompson, 2016). Yet, while these analyses did not unequivocally support a sex difference in the effect of obesity on renal failure, they rendered some support in favour of sex differences in the effect of obesity on risks of COPD and T2D.

Another potential issue for the analyses involving WHRadjBMI is collider bias (Davey Smith, Paternoster and Relton, 2017). However, Pulit *et al.* (2019) found limited evidence of collider bias in the GWAS where I obtained my WHRadjBMI instruments. To ascertain robustness, I also performed analyses investigating the effect of WHR on the cardiometabolic risk factors and diseases. The WHR and WHRadjBMI associations with disease endpoints were largely consistent, wherefore collider bias is unlikely to be severely biasing my results.

Most previous MR studies have only conducted sex-combined analyses. However, my results indicate that obesity traits may only increase risk of disease in one sex, or to different degrees in men compared to women. For example, there was clear evidence of WHRadjBMI impacting risk of chronic renal failure in men, but there was no evidence of an effect in women. Thus, in only conducting sex-combined MR analyses, we risk missing effects that are only present in one sex, or come to the wrong conclusion that a risk factor affects disease risk in both men and women.

2.4.9. Conclusion

In my first thesis chapter, I have carefully evaluated different approaches to create sex-specific GRSs of obesity and investigated the effect of obesity on the leading causes of death from non-communicable diseases. I found that use of sex-combined weights in instruments sometimes under/overestimated the effect of the instrument on the exposure in each sex, and that instrument strength could be improved by sex-specific weighting of instruments. In addition, I found that

obesity traits may increase risks of COPD, renal disease, and T2D to different degrees in men and women. This emphasises the importance of considering sex-specific effects when determining the consequences of obesity.

Chapter 3. Sex-specific effects of proteins on obesity and cardiometabolic disease

3.1. Introduction

3.1.1. The promise of proteomics

Several large-scale studies have performed GWASs to identify pQTLs and/or leveraged pQTLs to tease out relationships between proteins and diseases (Sun *et al.*, 2018; Yao *et al.*, 2018; Chong *et al.*, 2019; Ritchie *et al.*, 2019; Bretherick *et al.*, 2020; Folkersen *et al.*, 2020; Zheng *et al.*, 2020). In recent years, the sample sizes, number of investigated proteins, and identified pQTLs in these studies have increased dramatically (Sun *et al.*, 2018; Yao *et al.*, 2018; Chong *et al.*, 2019; Ritchie *et al.*, 2019; Bretherick *et al.*, 2020; Folkersen *et al.*, 2020; Zheng *et al.*, 2020). In 2017, Suhre *et al.* (2017) identified 539 pQTLs for 284 proteins in a total sample of 1,338 participants. The next year, Sun *et al.* (2018) identified 1,927 pQTLs for 1,478 proteins in a sample of 3,301 participants. Only two years later, a pQTL study in over 30,000 participants was published (Folkersen *et al.*, 2020). These, and other studies, have leveraged pQTLs to explore potential causal associations between proteins and diseases (Suhre *et al.*, 2017; Sun *et al.*, 2018; Yao *et al.*, 2018; Chong *et al.*, 2019; Ritchie *et al.*, 2019; Bretherick *et al.*, 2020; Folkersen *et al.*, 2020; Zheng *et al.*, 2020). Many of these studies have suggested new disease biomarkers, drug target candidates, and/or highlighted drugs that may be repurposed for new therapeutic indications (Sun *et al.*, 2018; Yao *et al.*, 2018; Chong *et al.*, 2019; Ritchie *et al.*, 2019; Bretherick *et al.*, 2020; Folkersen *et al.*, 2020; Gudmundsdottir *et al.*, 2020; Zheng *et al.*, 2020). For example, Folkersen *et al.* (2020) replicated several previous findings (such as a link between interleukin-6 and CAD), suggested repurposing opportunities for drugs targeting seven proteins, and highlighted eleven proteins as potential new drug targets.

3.1.2. The absence of sex-stratified genetic association studies of proteins

In spite of the increase in the scale of genetic association studies of proteins, the large-scale studies conducted to date have not explored sex-specific effects (Sun *et al.*, 2018; Yao *et al.*, 2018; Chong *et al.*, 2019; Ritchie *et al.*, 2019; Bretherick *et al.*, 2020; Folkersen *et al.*, 2020; Zheng *et al.*, 2020). Yet, a study by Oliva *et al.* (2020), using data from the Genotype-Tissue Expression project (GTEx), found that 37% of genes have sex-differential expression in at least one tissue and identified hundreds of eQTLs with sex-genotype interactions. The same study also reported that sex-differential gene regulation contributes to sex-differences in traits and diseases (Oliva *et al.*, 2020). Given these results, it seems plausible that there may also be sex differences in the regulation of protein levels and in the effects of proteins on diseases. However, up until now there has been no large-scale study that has been explored if pQTLs have the same effects in men and women, and if any relationships between proteins and cardiometabolic disease differ between the sexes.

3.1.3. Aims

In this chapter, my main research aim was to discover potential sex-specific protein biomarkers and drug targets for obesity and leading cardiometabolic diseases (Fig 3.1). To investigate this question further, I had two primary research objectives:

I. To perform sex-stratified GWASs of proteins to identify potential sex-specific pQTLs.

II. To investigate sex-specific associations of proteins on obesity, CAD and T2D using MR.

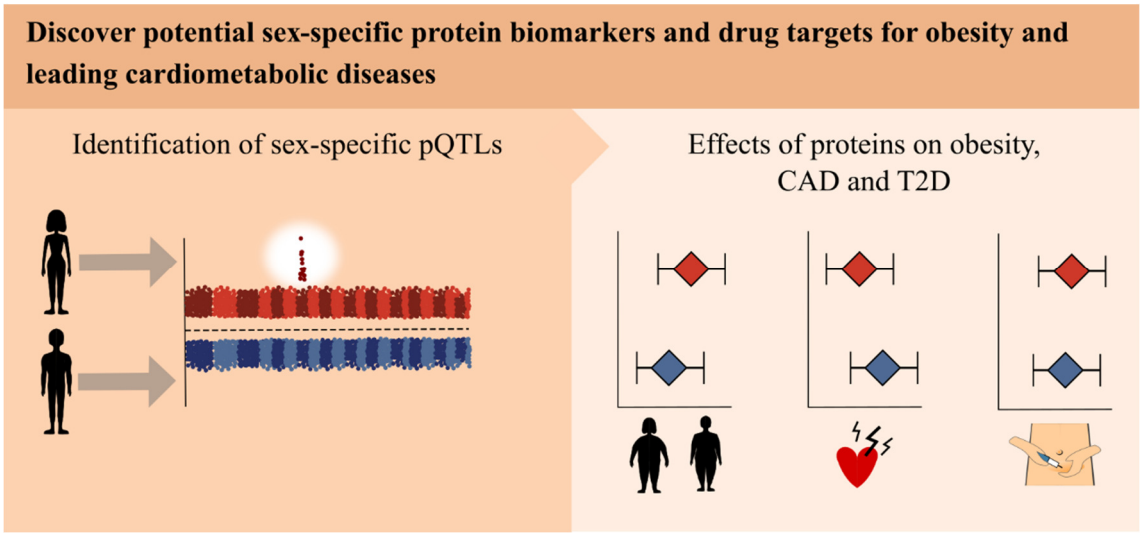


Fig 3.1. Graphical abstract of the primary research aim and research objectives for Chapter 3. My main research aim of Chapter 3 was to discover potential sex-specific protein biomarkers and drug targets for obesity and cardiometabolic disease. To this end, I had two research objectives; to identify sex-specific pQTLs and to determine the sex-specific associations of proteins on obesity, CAD, and T2D using MR.

3.2. Methods

3.2.1. Study overview

I first investigated overarching patterns of observational correlations between proteins in men versus women (Fisher, 1915, 1921; Osborne, 2017c). I proceeded to conduct sex-stratified GWASs to identify sex-specific pQTLs, before assessing the evidence of these pQTLs having sex-differential associations with protein levels (Randall *et al.*, 2013; Winkler *et al.*, 2015). Thereafter, I performed sex-stratified MR analyses to investigate the effects of proteins on obesity, T2D and CAD in men and women, and compared the MR estimates between the sexes (Cochran, 1954; Davey Smith and Ebrahim, 2003). Finally, I followed up with additional post-hoc analyses to explore my findings further and sensitivity analyses to explore the robustness of my results. Given the range of different methods that have been used in this chapter, I have added a subscript with the method name after the P-values for clarity.

3.2.2. Datasets

3.2.2.1. INTERVAL

3.2.2.1.1. Sample quality control

I received post-quality control individual-level data for the INTERVAL study from the authors of Sun *et al.* (2018). The entire INTERVAL study is a blood donor cohort with approximately 50,000 genotyped participants, recruited between 2012-2017 (Astle *et al.*, 2016; Di Angelantonio *et al.*, 2017). Since the study only included blood donors, participants with a previous history of major disease (such as myocardial infarction and stroke) and current infection at sampling time were excluded (Di Angelantonio *et al.*, 2017; Sun *et al.*, 2018). In the study by Sun *et al.* (2018), two non-overlapping sub-cohorts of the INTERVAL study had been randomly extracted for blood protein quantification. Sun *et al.* (2018) excluded any samples that were duplicated, had low genotype call rate, were of non-European ancestry (by PC analysis), extreme heterozygosity, or with sex mis-match. Only one sample in each pair related to a second degree or higher (defined as proportion of identity-by-descent >0.187) was included in the data (Sun *et al.*, 2018). After these genetic quality control procedures, the larger cohort included 1,202 women and 1,279 men and the smaller cohort included 414 women and 406 men (Table 3.1) (Sun *et al.*, 2018).

Table 3.1. Participant characteristics for INTERVAL.

Characteristic	Men	Women
N samples large cohort	1,279	1,202
Mean age (SD) large cohort	45.5 (14.5)	41.5 (13.9)
N samples small cohort	406	414
Mean age (SD) small cohort	46.8 (14.1)	41.4 (13.8)
N samples total	1,685	1,616
Mean age (SD) total	45.8 (14.4)	41.4 (13.8)

3.2.2.1.2. Genotype quality control

Genotyping had been performed using the Affymetrix Axiom UK Biobank genotyping array (Astle *et al.*, 2016; Di Angelantonio *et al.*, 2017; Sun *et al.*, 2018). A subset of 654,966 SNPs of high quality were used for imputation, using a combined version of UK10K and phase 3 of 1000

Genomes as a reference panel, resulting in a total of 10,572,788 imputed SNPs (Auton *et al.*, 2015; J. Huang *et al.*, 2015; The UK10K Consortium, 2015; Astle *et al.*, 2016; Di Angelantonio *et al.*, 2017; Sun *et al.*, 2018).

I performed post-imputation quality control, and excluded all SNPs that were indels, non-biallelic or duplicated. I further excluded SNPs with an imputation info score <0.7 or with HWE $P<5\times10^{-6}$, as in Sun *et al.* (2018). Finally, I also excluded SNPs with a minor allele count <20, computed separately in each sex and cohort, as the function used for computing the association statistics (--glm in PLINK) gives poorly calibrated results when the minor allele count is below 20 (Purcell *et al.*, 2007; Chang, 2021). In total, 7,172,280 genetic variants in women and 7,145,707 genetic variants in men passed quality control in both cohorts, with a total of 6,966,182 SNPs passing quality control in both sexes.

3.2.2.1.3. Aptamer quality control

Protein levels had been quantified in each cohort using the SOMAscan assay from SomaLogic (Fig 3.2) (SomaLogic, 2017; Sun *et al.*, 2018). The assay uses modified oligonucleotides marked with fluorophores – aptamers – that bind to specific protein epitopes (Gold *et al.*, 2010; SomaLogic, 2017; Sun *et al.*, 2018). Protein levels are then quantified using a DNA microarray (Gold *et al.*, 2010; SomaLogic, 2017). These measurements are then adjusted for systematic technical variability, such as between-plate variation, with the final protein levels given in relative fluorescent units (RFU) (Gold *et al.*, 2010; SomaLogic, 2017).

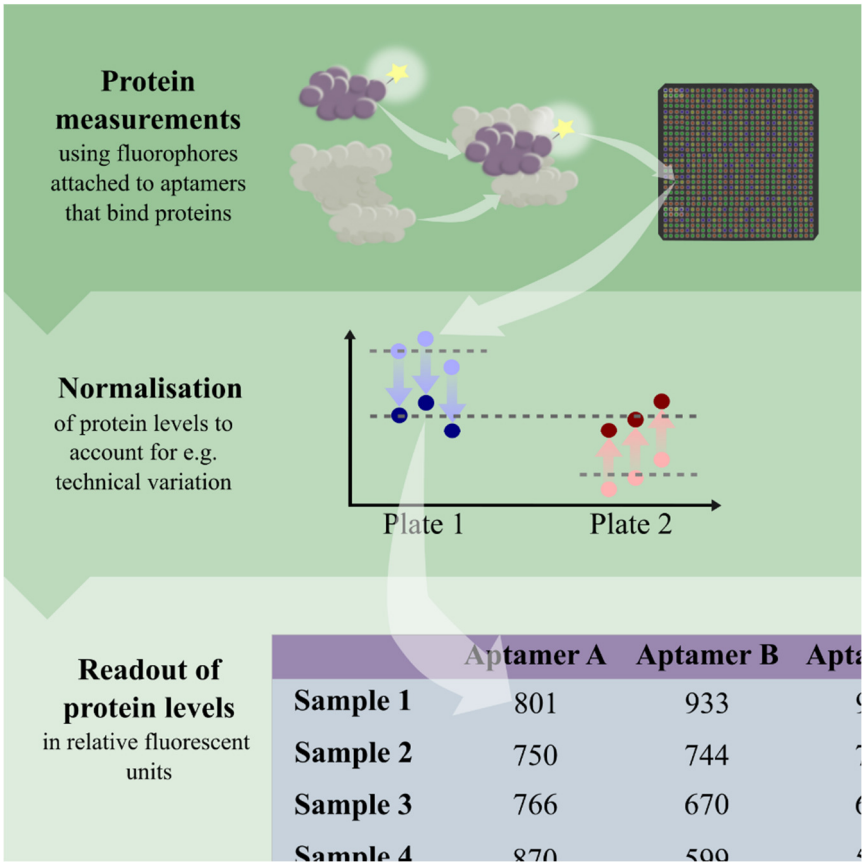


Figure 3.2. Simplified illustration of the SomaLogic protein quantification assay and normalisation procedures. Protein levels are quantified using fluorophore-marked aptamers with affinity to specific protein epitopes on a microarray (Gold *et al.*, 2010; SomaLogic, 2017). Thereafter, the protein levels are normalised to account for technical variations, before the protein levels for each sample is given in RFUs (Gold *et al.*, 2010; SomaLogic, 2017).

The SOMAscan assay version used in Sun *et al.* (2018) measured the concentrations of up to 4,034 aptamers (3,622 distinct proteins or protein complexes). However, Sun *et al.* (2018) excluded aptamers with a coefficient of variation >20% in either cohort and those targeting non-human proteins. After these quality control measures, 3,283 aptamers targeting 2,994 distinct proteins or protein complexes remained in the dataset provided to me (Sun *et al.*, 2018).

I made further exclusions of aptamers that failed various quality control measures (Fig 3.3). The aptamer LOC652493.6561.77.3 did not have a UniProt identifier and had been deprecated according to the National Center for Biotechnology Information (NCBI; see https://www.ncbi.nlm.nih.gov/gene/652493) and was therefore excluded (Brown *et al.*, 2015). Sun *et al.* (2018) had also performed cross-reactivity tests for a subset of the aptamers. To reduce the

risk of non-specific binding or poor aptamer quality affecting my results, I excluded all aptamers with any evidence of non-specific binding in these tests (Sun *et al.*, 2018). I also excluded all aptamers for which SomaLogic had any suspicion of poor quality or non-specific binding (2020, e-mail; January 30). Another study using the same INTERVAL data, by Ritchie *et al.* (2019), had also conducted additional quality control analyses of the aptamers, including specificity tests, and I excluded all aptamers that failed their quality control as well. Finally, aptamer information was extracted from Sun *et al.* (2018) and Ritchie *et al.* (2019) as needed. For the gene *IGK*, I based the gene position on the positive strand (as gene positions for both the positive and negative strand were provided for this gene) (Sun *et al.*, 2018; Ritchie *et al.*, 2019). After these quality control measures, a total of 2,840 aptamers measuring 2,623 unique proteins were taken forward for analysis.

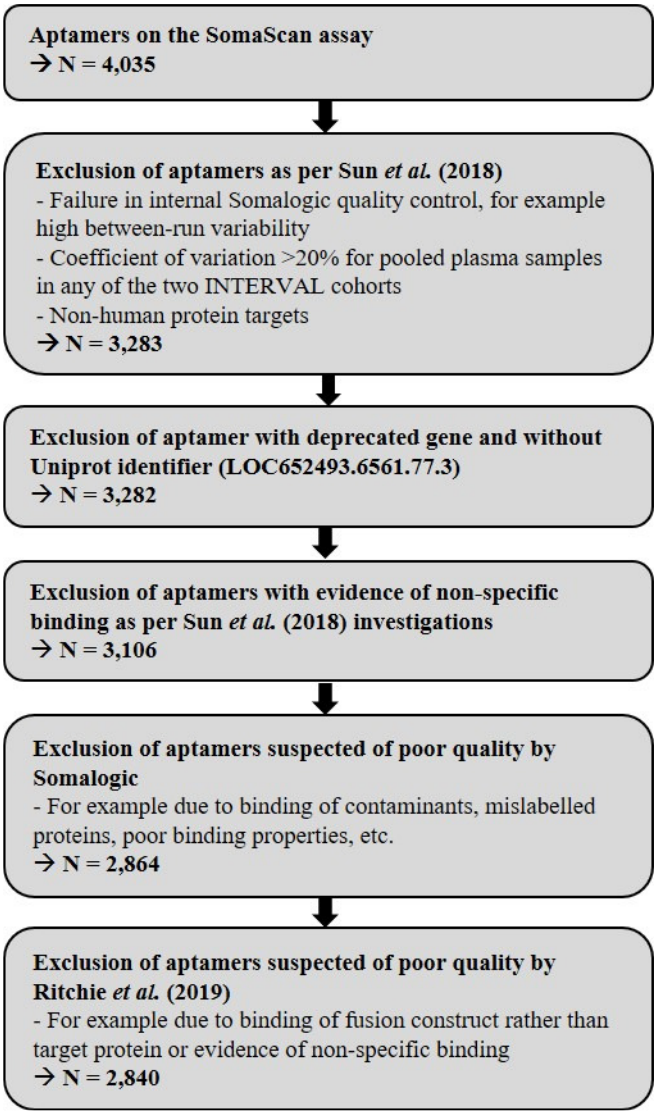


Figure 3.3. Aptamer quality control flowchart. Aptamers with any evidence of non-specific aptamer binding or poor aptamer quality were excluded from all analyses.

3.2.2.1.4. Transformation of aptamer measurements

Following a similar procedure as in Sun *et al.* (2018), I natural-log transformed the aptamer measurements. I then conducted a linear regression analysis of the aptamers adjusting for age, sample processing time (binary; ≤1 day or >1day), and three PCs (computed by Sun *et al.* (2018)), performed separately in each sex and cohort (Sun *et al.*, 2018). I then normalised the residuals from this regression using RINT, also conducted separately in each sex and cohort (Bliss, 1967; Osborne, 2017c; Sun *et al.*, 2018).

3.2.2.2. UK Biobank

3.2.2.2.1. Sample quality control

The UKBB data had been processed centrally as described in section 2.2.2 (Sudlow *et al.*, 2015; Bycroft *et al.*, 2018). For the analyses in this chapter, I performed sample quality control using the same criteria as described in section 2.2.2, but restricted the sample to only contain those of "White British ancestry". This resulted in a final sample size of 174,637 men and 202,554 women (Table 3.2).

Table 3.2. Participant characteristics for UKBB.

Characteristic	Men	Women
N samples	174,637	202,554
Mean age (SD)	57.2 (8.0)	56.7 (7.9)
Mean BMI (SD)	27.8 (4.2)	27.0 (5.1)
Mean WHR (SD)	0.9 (0.1)	0.8 (0.1)
Arrayed on UKBL (%)	20,862 (11.9%)	20,437 (10.1%)
CAD cases (%)	22,046 (12.6%)	10,323 (5.1%)
T2D, probable and possible cases (%)	11,103 (6.4%)	6,186 (3.1%)
T2D, probable cases only (%)	9,675 (5.5%)	5,355 (2.6%)

The UKBB has a Research Tissue Bank approval, and all participants gave informed consent (Sudlow *et al.*, 2015; Bycroft *et al.*, 2018). The work presented in this chapter was performed under UKBB study application ID 11867.

3.2.2.2.2. UK Biobank genotype quality control

I excluded all SNPs with an imputation info score $\leqslant$0.3, HWE $P<1\times10^{-6}$, minor allele frequency <1%, indels, or that were non-biallelic or duplicates (Bycroft *et al.*, 2018; Yengo *et al.*, 2018).

3.2.2.3. Summary statistics for obesity traits

Sex-specific summary data for BMI, WHR, and WHRadjBMI were taken from the GIANT and UKBB meta-analyses by Pulit *et al.* (2019), performed in participants of European ancestry only and described in detail in section 2.2.4.1. I excluded SNPs that were indels, duplicated, or which had SNP information that was inconsistent with data from NCBI (using the file "All_20180423.vcf.gz" from

ftp://ftp.ncbi.nih.gov/snp/organisms/human_9606_b151_GRCh37p13/VCF/ (Sherry *et al.*, 2001)). Only SNPs that passed quality control in both men and women were kept for downstream analyses.

3.2.2.4. LD reference panels and genome build

I constructed one LD-reference panel from the INTERVAL dataset and one LD-reference panel from the UKBB dataset, which were each used for the analyses involving the respective dataset. For INTERVAL, I converted the genotype dosages to hard-call genotypes using a hard-call threshold of 0.1 using PLINK (Purcell *et al.*, 2007). For the UKBB LD-reference panel, 1% (N=3,384) of the participants were extracted from the main UKBB dataset, after exclusion of all participants related to a third degree or closer (kinship coefficient>0.0442) (Manichaikul *et al.*, 2010; Bycroft *et al.*, 2018). The quality-controlled genotype probabilities were then converted to hard-call genotypes using a threshold of 0.1, and I further excluded variants with a missingness >0.05 (Yengo *et al.*, 2018).

Genome build GRCh 37 was used throughout to denote genomic positions and to calculate distances.

3.2.3. Observational correlations between pairs of aptamers

3.2.3.1. Correlation analyses

I first explored aptamer-aptamer correlations using observational data, before investigating if the correlations differed between men and women. To this end, linear regression analyses of the normalised protein values were conducted in each sex and cohort (Osborne, 2017c). The two cohorts were then meta-analysed using an IVW fixed effects model using the metafor R package (Viechtbauer, 2010). I then further meta-analysed the men- and women-specific correlations to obtain sex-combined aptamer-aptamer correlations for comparison (Viechtbauer, 2010). The multiple testing threshold was set to the P-value for the observational correlations ($P_{correlation}$) threshold corresponding to <1% FDR using all P-values from all aptamer pairs in the men, women, and sex-combined analyses (Benjamini and Hochberg, 1995).

3.2.3.2. Assessing sex differences in aptamer-aptamer correlations

Aptamer pairs that correlated $P_{correlation}$<1% FDR in either men or women were assessed for a sex difference between the sex-specific correlation *r*-values. Correlations were converted to Z-scores using a Fisher z' transformation (Fisher, 1915, 1921) and code from Bishara and Hittner (2017). Sex differences in the aptamer-aptamer correlations were then computed as:

$$Z_{correlation-sexdiff} = \frac{Z_{men} - Z_{women}}{\sqrt{\frac{1}{n_{men} - 3} + \frac{1}{n_{women} - 3}}}$$

where Z_{men} and Z_{women} represent the Fisher z' transformed correlations, and the sample sizes in each sex are given as n_{men} and n_{women} (Fisher, 1915, 1921). The $Z_{correlation-sexdiff}$ was then converted to a P-value for a sex difference in the observational correlations ($P_{correlation-sexdiff}$) using a two-sided transformation (Fisher, 1915, 1921).

3.2.4. Post-hoc functional enrichment analyses

A substantial proportion of the sex-differential aptamer-aptamer correlations seemed to be explained by a small subset of aptamers. Thus, I decided post-hoc to investigate this subset further with a functional gene set enrichment analysis using the g:GOSt software (Raudvere *et al.*, 2019). In brief, g:GOSt uses information from resources such as REACTOME (Fabregat *et al.*, 2018), KEGG (Kanehisa *et al.*, 2019), and Ensembl (Yates *et al.*, 2020) to investigate enrichment of biological processes, pathways, and regulatory motifs in a gene set (Raudvere *et al.*, 2019). For the gene set, I included all the genes coding for any protein targeted by the aptamers that were sex-differentially correlated with >5% of all other aptamers. This threshold was chosen post-hoc so as to capture an apparent spike in the data (see Results and Fig 3.8). The gene background input was set to all the genes coding for any of the proteins and protein complexes targeted by any aptamer in this study.

3.2.5. Sex-specific genome-wide association studies of protein levels

GWASs of the 2,840 aptamers were performed separately in each sex and cohort using genotype dosages and the --glm command in PLINK version 2 (Purcell *et al.*, 2007). The sex-stratified results from the two cohorts were then meta-analysed using the metal software and an IVW model (Willer, Li and Abecasis, 2010). To correct for multiple testing, the statistical significance threshold was to the P_{GWAS}-value corresponding to 1% FDR, computed from the P-values of all SNPs passing quality control in both men and women and all aptamers (Benjamini and Hochberg, 1995). Given the large number of P_{GWAS}-values, I modified code originally created by McBryan and Leonardi (McBryan, 2013) to perform the FDR-conversion (Benjamini and Hochberg, 1995). For comparison, the men- and women-stratified GWASs were then meta-analysed to obtain sex-combined effect estimates, using metal and an IVW model (Willer, Li and Abecasis, 2010).

To evaluate potential sex differences between the effect estimates for men and women, $Z_{sexdiff}$-values were computed using:

$$Z_{sexdiff} = \frac{\beta_{women} - \beta_{men}}{\sqrt{SE_{women}^2 \times SE_{men}^2}}$$

using EasyStrata (Randall *et al.*, 2013; Winkler *et al.*, 2015), and then converted to $P_{sexdiff}$-values using a two-sided transformation. For computational reasons, this formula was applied in the main GWAS pipeline. I then extracted the pQTLs in each sex that were associated with each protein at P_{GWAS}<1% FDR and with consistent effect directions at P_{GWAS}<0.05 in both cohorts for that sex. The multiple testing threshold was set to the $P_{sexdiff}$-value corresponding to 1% FDR computed amongst these SNPs (performed once for all SNP-aptamer pairs) (Benjamini and Hochberg, 1995). Finally, a set of independent SNPs for each aptamer was identified using clumping, an r^2 threshold of 0.01, and a 5 Mb window size in PLINK (Purcell *et al.*, 2007).

3.2.6. Sex-specific genetic aptamer-aptamer correlations

I used LDSC to compute sex-specific genetic correlations between aptamer pairs (B. Bulik-Sullivan *et al.*, 2015; B. K. Bulik-Sullivan *et al.*, 2015). The sex-specific GWAS summary statistics data were used, after further exclusion of SNPs with an imputation info score ≤0.9 and only keeping SNPs that passed quality control in both sexes in both cohorts (Bulik-Sullivan and Finucane, 2017b). Publicly available LD scores for Europeans were used (B. K. Bulik-Sullivan *et al.*, 2015; Bulik-Sullivan and Finucane, 2017b). Only aptamers with evidence of a polygenic genetic architecture (defined as mean $chi^2 \geq 1.02$ in either sex) were investigated, as LDSC is based on a polygenic model (B. Bulik-Sullivan *et al.*, 2015; B. K. Bulik-Sullivan *et al.*, 2015; Bulik-Sullivan and Finucane, 2017a). Sex-stratified genetic correlations between aptamer pairs were then computed using the --rg command using LDSC in python (B. Bulik-Sullivan *et al.*, 2015; B. K. Bulik-Sullivan *et al.*, 2015; Python Software Fondation, 2020). The multiple testing threshold was set to the genetic correlation P-value (P_{LDSC}) corresponding to <1% FDR using the P_{LDSC}-values from all aptamer pairs where genetic correlation could be computed (Benjamini and Hochberg, 1995).

3.2.7. Sex-stratified genome-wide association studies of type 2 diabetes and coronary artery disease

There was no publicly available sex-stratified GWAS summary statistics for T2D or CAD performed in a large European sample available for use in the MR analyses. I therefore performed sex-stratified GWASs of these diseases in the UKBB (Sudlow *et al.*, 2015).

3.2.7.1. Disease case and control criteria

Cases and controls for T2D were extracted using code kindly provided by the authors of Udler *et al.* (2018). In brief, the code is an expansion on a previous diabetes case algorithm developed by Eastwood *et al.* (2016). The code will classify participants as having “probable” T2D, “possible” T2D, other kinds of diabetes, or conclude that the individual is unlikely to have diabetes (Eastwood

et al., 2016; Udler *et al.*, 2018). Information from many different sources is used, for example data on self-reported diseases, medications, diabetes age-of-onset, and information from an interview with a nurse (Eastwood *et al.*, 2016; Udler *et al.*, 2018). For my study, participants that the algorithm defined as having "possible" or "probable" T2D were classified as T2D cases (Eastwood *et al.*, 2016; Udler *et al.*, 2018). Participants that the algorithm defined as unlikely to have diabetes were classified as controls, and participants that were listed as potentially having other diabetes types were excluded (Eastwood *et al.*, 2016; Udler *et al.*, 2018)..

For CAD, the same case and control definition as described in section 2.2.5 was used, which was originally based on a GWAS of CAD by Nelson *et al.* (2017).

3.2.7.2. Genome-wide association studies of coronary artery disease and type 2 diabetes

Sex-specific and sex-combined GWASs for CAD and T2D were then computed with a logistic regression and the --glm function in PLINK (Purcell *et al.*, 2007). Adjustments were made for age, array type, 20 PCs, and additionally sex in the sex-combined analyses (Emdin *et al.*, 2017; Young *et al.*, 2019).

3.2.8. Mendelian randomisation analyses of proteins on cardiometabolic traits

3.2.8.1. Protein instruments

Instruments for each aptamer were constructed using the following approaches in each sex (the index approach was also conducted sex-combined) (Fig 3.4):

a) **Index:** All sex-specific index pQTLs for the aptamer across the genome, thereby maximising the number of independent SNPs in the instrument (similar to the pan-pQTL approach by Folkersen *et al.* (2020)).
b) **Cis:** All cis-pQTLs (within 500 kb of an aptamer gene) which also have consistent effects on the protein using any other available aptamers (consistent effect direction and P_{GWAS}<1% FDR) (Benjamini and Hochberg, 1995). This approach was inspired by the cis-

pQTL approach by Folkersen *et al.* (2020). The cis-instrument has the benefit of generally not including the highly pleiotropic trans-pQTLs highlighted by Sun *et al.* (2018). In addition, the cis-instrument excluded pQTLs with evidence of poor consistency between aptamers targeting the same protein. However, the cis-instrument has a potentially higher risk of being more strongly affected by changed aptamer binding if a pQTL is in LD with a coding variant for the targeted protein (Sun *et al.*, 2018).

c) **Men/women pQTLs:** The index pQTLs identified in the other sex (for example, the men index pQTLs), but weighted using the effect estimates from the sex under investigation (for example, women effect estimates). This instrument was created to allow a fair comparison between the sexes, as otherwise sexual heterogeneity in the MR estimates potentially could be solely because of the SNPs that were included in the men versus the women instruments.

Each instrument was then further pruned by excluding the SNP with the highest P_{GWAS}-value in each pair correlated at $r^2 \geq 0.01$ to reduce the possibility of long-distance LD affecting the results.

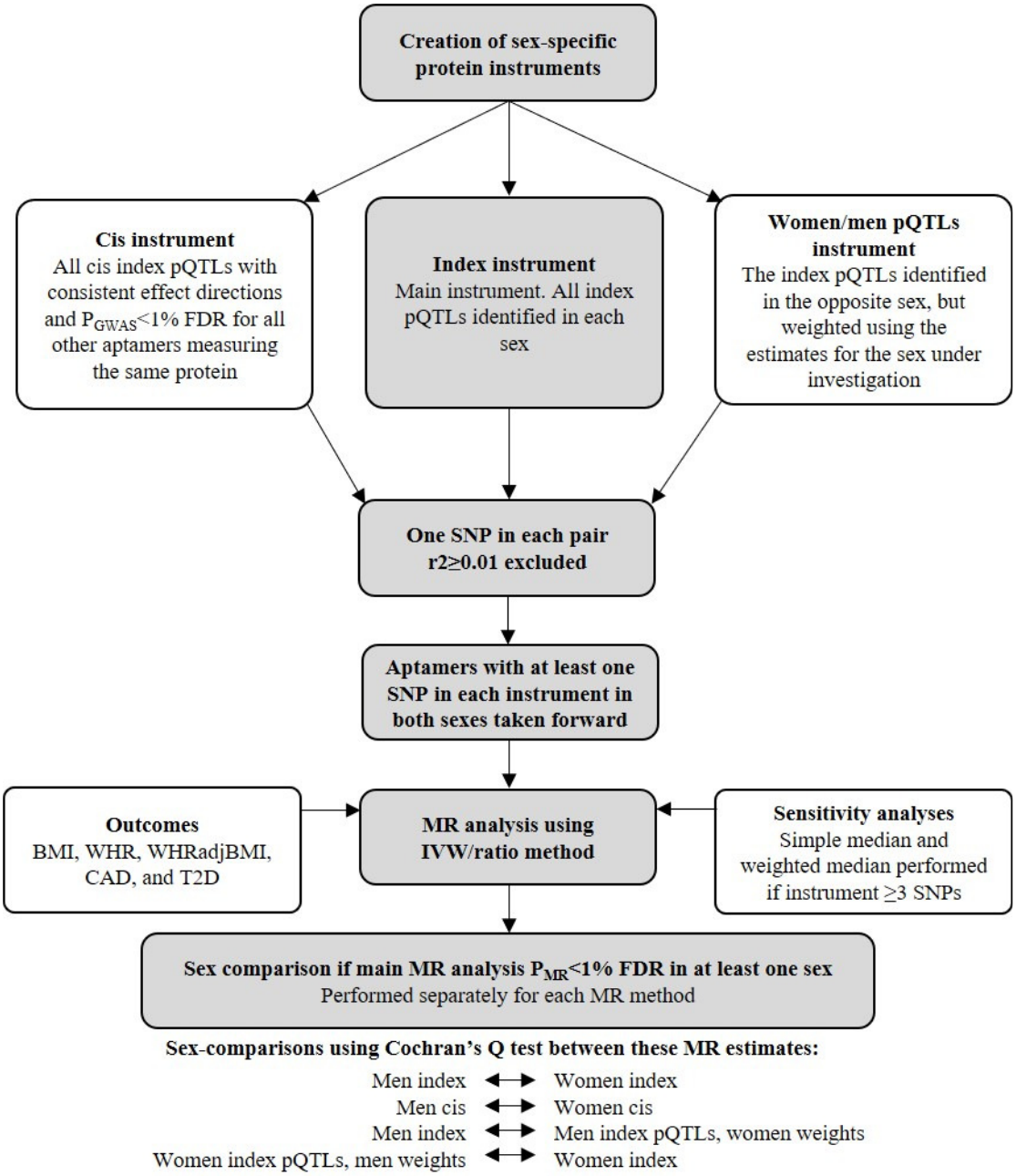


Figure 3.4. Flowchart of the instrument creation protocol and MR analyses. Three different types of sex-specific instruments were created for each aptamer measuring a protein. Aptamers with at least one SNP in each instrument in both men and women were taken forward for to estimate the sex-specific effect of the protein on the obesity traits and cardiometabolic diseases. If the main protein-outcome MR analysis, using the index instrument and the IVW or ratio method (Burgess and Thompson, 2015), surpassed the MR P-value (P_{MR}) multiple testing threshold (P_{MR}<1% FDR) (Benjamini and Hochberg, 1995), sexual heterogeneity between the effect estimates in men and women was assessed using a Cochran's Q test (Cochran, 1954).

In addition, sex-combined index instruments were created for comparison, using the same approach as for the sex-specific index instruments.

Palindromic SNPs, i.e., those with A/T or C/G alleles, were excluded from the MR analyses, unless the minor allele frequency was below 0.4 in both datasets and the minor allele was the same in both datasets (Burgess and Thompson, 2015; Burgess, Davey Smith, *et al.*, 2020). If a SNP was missing in the outcome dataset, the proxy with the highest r^2 and closest position was used instead (with a minimum $r^2 \geq 0.8$ and within 250 kb) (Burgess *et al.*, 2015; Adams and Neuhausen, 2019; Burgess, Davey Smith, *et al.*, 2020). If needed, mapping between datasets was performed using additional SNP information from NCBI (using the files "All_20180423.vcf.gz" from ftp://ftp.ncbi.nih.gov/snp/organisms/human_9606_b151_GRCh37p13/VCF/ and "RsMergeArch.180920.txt.gz" from ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606/database/organism_data/) (Sherry *et al.*, 2001).

3.2.8.2. Mendelian randomisation analysis of the effects of proteins on obesity and cardiometabolic disease

Only aptamers with at least one SNP in each instrument type in both men and women were taken forward for MR analysis. The IVW method (or the ratio method in the case of a single-SNP instrument) were used as the main method (Fig 3.4) (Burgess, Butterworth and Thompson, 2013; Burgess *et al.*, 2015). If an instrument included ≥3 SNPs, I additionally performed simple and weighted median MR analysis as a sensitivity analysis (Bowden *et al.*, 2016). All MR analyses were conducted using the MendelianRandomization R package (Yavorska and Burgess, 2017). The multiple-testing threshold was set to P_{MR}<1% FDR, computed in all main index instrument IVW/ratio method MR analyses in men and women (Benjamini and Hochberg, 1995).

If a protein had an impact on a cardiometabolic trait or disease in at least one sex in the main MR analysis, heterogeneity between the sex-specific effect estimates was assessed using Cochran's Q test (Cochran, 1954). If the P-value for the sex-comparison between the MR estimates (P_{het}) surpassed a <1% FDR threshold (computed using the P_{het}-values from all the comparisons of the main MRs), the protein was considered to have sexually heterogenous effects on the outcome

(Benjamini and Hochberg, 1995). As a sensitivity analysis, sex-comparisons using Cochran's Q test (Cochran, 1954) were also conducted:

a) Between the sex-specific effect estimates using the cis instruments in each sex.

b) Between the sex-specific effect estimates using the index instrument in one sex and the Men/women pQTL instrument in the other sex. In this approach, effect estimates were compared from MR estimates computed using the same SNPs (but weighted by their sex-specific effect estimates).

These sex-comparisons were also made for the MR estimates computed using the simple and weighted median methods (Bowden *et al.*, 2016).

To reduce the risk of false positive results, I devised a set of criteria to select "high confidence" MR results with robust support from the sensitivity analyses (Fig 3.5).

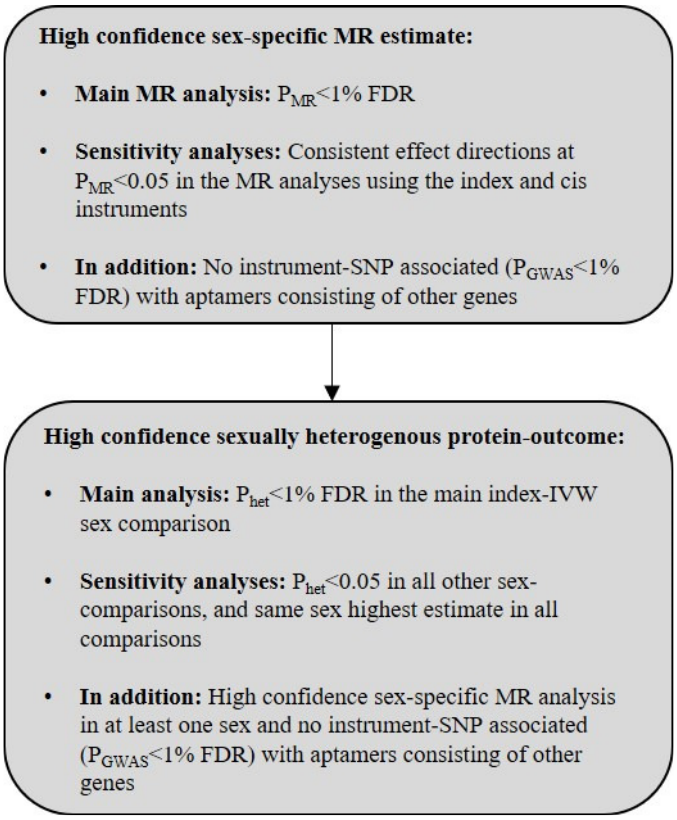


Figure 3.5. Criteria to select "high-confidence" MR results. To reduce the risk of false positive results, I defined criteria to identify protein-outcome combinations with consistent results in the sensitivity analyses.

3.2.9. Sensitivity analyses

3.2.9.1. Reconducting the observational aptamer-aptamer correlations using a different protein transformation approach

In the main analysis, aptamer-aptamer correlations were computed using protein levels that had been transformed through RINT of the residuals after a linear regression of the natural log-transformed proteins in RFUs (Bliss, 1967; Osborne, 2017c; Sun *et al.*, 2018). As a sensitivity analysis, I also reconducted the correlation analyses using another protein transformation approach. In this approach, I first did a RINT of the proteins in RFUs (Bliss, 1967; Pain, Dudbridge and Ronald, 2018). I then conducted a linear regression with the same adjustments as in the main analysis, and did a second RINT of the residuals from this regression (Bliss, 1967; Osborne, 2017c; Pain, Dudbridge and Ronald, 2018). Thereafter, correlations were computed in each sex-strata using the same procedure as in the main analysis.

3.2.9.2. Investigations into the agreement between different aptamers measuring the same protein

As a sensitivity analysis, I decided to investigate the agreement between aptamers targeting the same protein. I first quantified the proportion of aptamer-aptamer combinations targeting the same protein target that were positively correlated in the observational data ($P_{correlation}$<1% FDR). Thereafter, I investigated the support for the index pQTLs that were targeting a protein measured by several aptamers. To this end, I quantified the proportion of pQTLs with consistent effect directions a) P_{GWAS}<0.05 and b) P_{GWAS}<1% FDR in all other aptamers measuring the same protein (in the subset of pQTLs that measured a protein targeted by several aptamers).

3.2.9.3. Post-hoc exploration of the evidence of a protein quantitative trait locus with opposite effect directions in men and women using phenotype data

The SNPs rs35185467 and rs17065165 represented a locus which seemed to have opposite effect directions on levels of angiopoietin-4 and ephrin-B3 in men versus women. Previous studies have indicated that both angiopoietin-4 and ephrin-B3 are connected to angiogenesis and blood pressure

(Yancopoulos *et al.*, 2000; Olsen *et al.*, 2006; Pfaff, Fiedler and Augustin, 2006; Wang *et al.*, 2016). I therefore investigated if these SNPs were associated with potentially related phenotypes, and if any such associations differed between men and women. To this end, I looked up the sex-specific SNP associations in publicly available sex-stratified GWASs of SBP, DBP and deep venous thrombosis (phenotypes "4080_irnt", "4080_irnt", and "20002_1094") (Neale lab, 2020).

3.2.9.4. Assessing the robustness of the results to protein transformation approach and adjustment protocol

3.2.9.4.1. Additional protein transformation approaches and adjustment protocols evaluated

As a sensitivity analysis, I reconducted the regressions between each index pQTL identified in the main protein GWASs with their respective aptamer using additional protein transformation approaches and adjustment protocols (Table 3.3).

Table 3.3. Sensitivity analysis of effect of protein transformation approach and adjustment protocol.

Protein transformation and adjustment protocol	Protein transformation	Adjusted age^2	Removal of outliers	Adjustments in GWAS instead of normalisation step	Same N men and women
Normalised	Sex-stratified RINT of residuals after regression on log-transformed proteins				
Normalised, adjusted age^2		X			
Normalised, no outliers			X		
Normalised, PCs adjusted in GWAS				PCs	
Rank-normalisation	Sex-stratified RINT of residuals after regression on RINTed proteins				
Rank-normalisation, PCs adjusted in GWAS				PCs	
Rank-normalisation, adjustments in GWAS				All covariates	
Normalised sex-jointly	Sex-jointly performed RINT of residuals after regression on log-transformed proteins, then GWAS performed sex-stratified				
Normalised sex-jointly, same N					X
Natural log	Natural log transformed only				
Natural log, no outliers			X		
Natural log, same N					X
Natural log, no outliers, same N			X		X

All analyses were performed with adjustments for age, blood processing time, and 3 PCs (Sun *et al.*, 2018). Removal of outliers denotes the removal of participants with log-transformed protein values >3 SD from the mean in that sex and cohort. Adjustments in GWAS instead of normalisation step denotes what covariates that were adjusted for in the GWAS step instead of in the protein normalisation step. Same N men and women denotes that the sample was randomly subset to have the same number of men and women in each cohort.

The sex-stratified RINT of the residuals after regression on log-transformed proteins was performed to account for differences in protein distributions in men and women, similar to previous GWASs where the phenotype distributions may differ between the sexes (Pulit *et al.*, 2019; Ruth *et al.*, 2020). The rank-normalisation approaches, where RINT was effectively conducted twice, was performed to reduce the risk of re-introduced associations with covariates that can occur in the case of a standard RINT (Pain, Dudbridge and Ronald, 2018). The normalised

sex-jointly and natural-log approaches were conducted so that the SNP estimates would be in the same absolute units in both sexes, albeit with a high risk of invalidating the linear regression assumptions of, for example, normally distributed residuals (Osborne, 2017a; Pirastu *et al.*, 2020). The adjustment for age^2 was conducted in case the protein-age relationship would not be linear for all proteins, similar to models for other traits (Pulit *et al.*, 2019). The removal of outliers was done to minimise the risk of the association being driven by outliers in the data (Osborne, 2017c). To examine the impact of *when* adjustment for covariates was made, I additionally tested different versions of this parameter as well, since previous studies have used different approaches (Pain, Dudbridge and Ronald, 2018; Sun *et al.*, 2018; Folkersen *et al.*, 2020). Finally, I also conducted the analyses using a random subset of the sex with the most samples. This was done to ensure a similar sample size in both cohorts, to minimise the risk of sex differences being due to the cohorts having slightly different sample sizes of men and women.

These different protein transformation approaches and adjustment protocols were all performed in the full sample as well as a <40 years old age-group and >60 years old age-group, in case for example menopause would cause sex differences to vary by age-strata (Rannevik *et al.*, 1995). As an extra safety precaution, these analyses were performed using another genetic association software (SNPtest, (Marchini *et al.*, 2007)) than that used in the main analysis. $P_{sexdiff}$-values were then computed between men- and women effect estimates from each protocol and age-group, using the same formula as in the main analysis (Randall *et al.*, 2013). The same multiple testing thresholds as in the main analysis were used to denote statistical significance.

3.2.9.4.2. Criteria to define high-confidence sex-differential protein quantitative trait loci

I then devised a set of criteria to select pQTLs with robust evidence of having sex-differential effect estimates in these sensitivity analyses:

a) $P_{sexdiff}$<1% FDR in all sensitivity analyses performed in the full sample, and with the same sex consistently having the highest effect estimate.

b) $P_{sexdiff}$<0.05 in all sensitivity analyses performed in either the <40 or the >60 age-group, and with the same sex consistently having the highest effect estimate.

c) For each sex that the pQTL was associated with the aptamer in, the pQTL is also associated with all other aptamers targeting the same protein (at an association threshold of P_{GWAS}<1% FDR and with consistent effect directions).

3.2.9.5. Probable type 2 diabetes cases only

In the main analysis, the sex-stratified GWASs and MRs were performed using all available (probable and possible) T2D cases from the diabetes case definition algorithm (Udler *et al.*, 2018). I also reconducted these analyses using only the T2D cases the algorithm had coded as having "probable" T2D to minimise the risk of my results being affected by misclassified diabetes type cases (Udler *et al.*, 2018).

3.2.9.6. Bidirectional Mendelian randomisation analyses

I performed bidirectional MR analyses for all protein-outcome pairs that surpassed the multiple testing threshold in the main IVW/ratio method MR analysis (using index instruments) (Benjamini and Hochberg, 1995).

3.2.9.6.1. Obesity and cardiometabolic disease instruments

I constructed sex-specific index instruments for BMI, WHR, WHRadjBMI, T2D and CAD. For these instruments, I only included SNPs that passed quality control in both men and women, and that were available in the UKBB LD reference panel. Independent sex-specific index SNPs were selected using the --clump command in PLINK (using a P_{GWAS}-threshold of 5×10^{-8}, r^2 threshold of 0.01, and using a window size of 5 Mb) (Purcell *et al.*, 2007).

3.2.9.6.2. Mendelian randomisation analyses

The bidirectional MR analyses were conducted in the same manner as the main analyses, with the exception that MR-Egger analysis was also performed if an instrument included ≥10 SNPs

(Bowden, Davey Smith and Burgess, 2015; Burgess and Thompson, 2015; Burgess *et al.*, 2015; Bowden *et al.*, 2016; Yavorska and Burgess, 2017).

3.2.9.7. Investigations into potential sex-differential participation bias

I conducted two sensitivity analyses to discover potential sex-differential participation bias, similar to a previous study by Pirastu *et al.* (2020). I first performed a GWAS of sex in each INTERVAL cohort using SNPtest (Marchini *et al.*, 2007), adjusting for age, blood processing time, and 3 PCs (Sun *et al.*, 2018). The results from the two cohorts were then meta-analysed using an IVW model in metal (Willer, Li and Abecasis, 2010).

In the first sensitivity analysis, I investigated if any SNPs were associated with sex at P_{GWAS}<5% FDR, either in a single cohort or in the meta-analysis (Benjamini and Hochberg, 1995). The FDR-conversions were conducted in a) the two cohorts jointly and b) in the meta-analysis separately.

As a second sensitivity analysis, I attempted to compute a heritability h^2 measure of sex using LDSC (B. K. Bulik-Sullivan *et al.*, 2015). In this analysis, all genetic variants passing quality control in both sexes and with an imputation info score >0.9 were included (Bulik-Sullivan and Finucane, 2017b). I used publicly available LD scores for Europeans (B. K. Bulik-Sullivan *et al.*, 2015; Bulik-Sullivan and Finucane, 2017b). Thereafter, I evaluated the mean chi^2-value to determine if the phenotype was suitable for LDSC, setting the chi^2-threshold for suitability to 1.02, as previously recommended (B. K. Bulik-Sullivan *et al.*, 2015; Bulik-Sullivan and Finucane, 2017a).

3.2.10. Data availability and software

3.2.10.1. Data availability

For data availability see Appendix Table 1.1.

3.2.10.2. Software

Analyses were done using PLINK versions 1.90b3 and v2.00a2LM (Purcell *et al.*, 2007), SNPtest version 2 (Howie, Donnelly and Marchini, 2009), metal (Willer, Li and Abecasis, 2010), QCTOOL version 2 (Band and Marchini, 2020), and LDSC version 1.0.1 (B. Bulik-Sullivan *et al.*, 2015; B. K. Bulik-Sullivan *et al.*, 2015). Additional analyses and plots were done using R versions 3.5.1 and 3.4.3 (R Core Team, 2014), bash version 4.1.2(2) and awk (Free Software Foundation, 2020), and python version 2.7.15 (Python Software Fondation, 2020). The following R packages were used; EasyStrata (Winkler *et al.*, 2015), data.table (Dowle and Srinivasan, 2020), gprofiler2 (Raudvere *et al.*, 2019), MendelianRandomization (Yavorska and Burgess, 2017), tidyr (Wickham and Henry, 2018), splitstackshape (Mahto, 2019), R.utils (Bengtsson, 2017), psych (Revelle, 2019), mada (Doebler, 2017), metafor (Viechtbauer, 2010), xlsx (Dragulescu and Arendt, 2018), ggdendro (de Vries and Ripley, 2016), and ggplot2 (Wickham, 2016).

3.3. Results

3.3.1. Wide-spread correlations between aptamers using observational data

In total, there were 4,031,380 unique aptamer pairs in each sex-strata. Of these, approximately 61% of all aptamer pairs in each sex were correlated $P_{correlation}$<1% FDR (corresponding to $P_{correlation}$≤0.006; 2,448,647 pairs in men and 2,458,043 pairs in women) (Benjamini and Hochberg, 1995). The fraction of aptamer pairs that were correlated ($P_{correlation}$<1% FDR) was slightly higher (71%, or 2,859,150 pairs) in the sex-combined meta-analysis (Fig 3.6) (Benjamini and Hochberg, 1995).

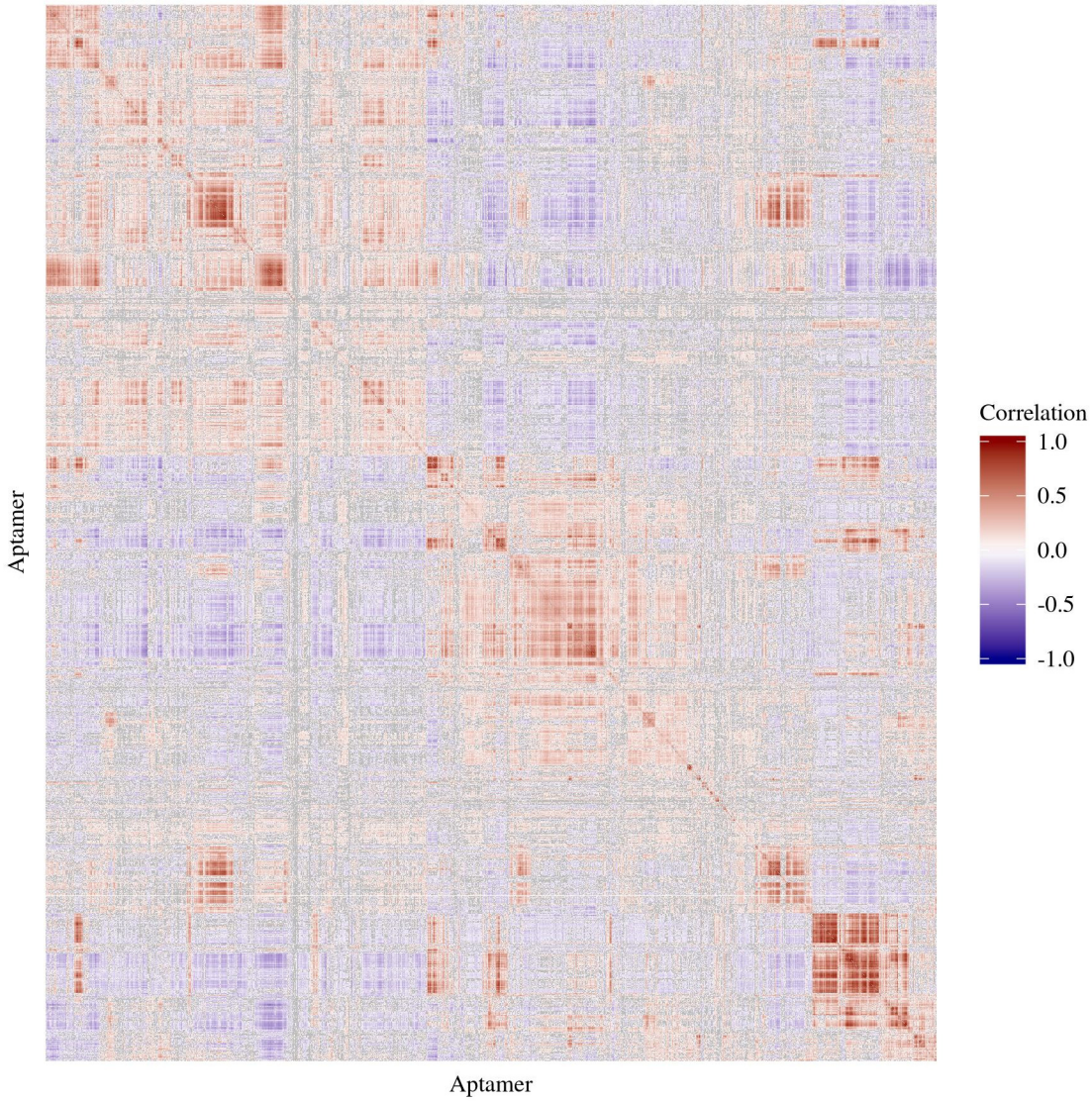


Figure 3.6. Observational correlations between all possible aptamer pairs in the sex-combined analysis. Only aptamer pairs correlated at $P_{correlation}$<1% FDR have their correlation *r* values colour-coded; non-significant aptamer-aptamer correlations are coded grey (Benjamini and Hochberg, 1995).

In general, there was a large agreement between the men- and women-stratified analyses. Of the aptamer pairs with $P_{correlation}$<1% FDR in men and/or women, only 0.76% (21,395) pairs had evidence of having sex-differential correlation *r* values ($P_{correlation\text{-}sexdiff}$<1% FDR, corresponding to $P_{correlation\text{-}sexdiff} \leq 7.6\times10^{-5}$) (Fig 3.7) (Benjamini and Hochberg, 1995).

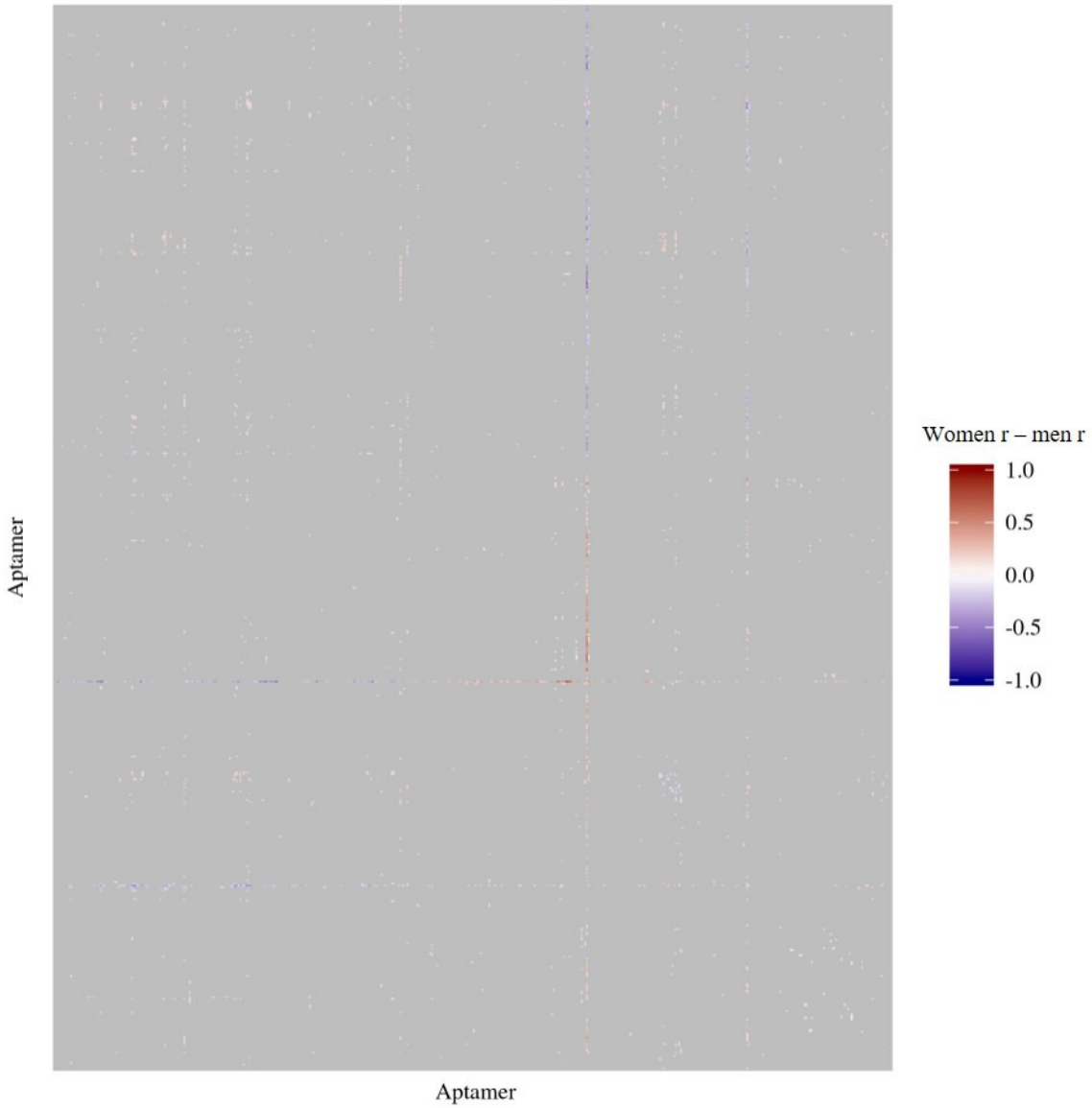


Figure 3.7. Sex-differentially correlating aptamer pairs. Only aptamer pairs sex-differentially correlated at $P_{correlation\text{-}sexdiff}$<1% FDR (Fisher, 1915, 1921; Benjamini and Hochberg, 1995) have the sex difference in the correlation *r* values colour-coded; aptamer-aptamer correlations that are not sex-differential are coded grey. Only aptamer pairs with $P_{correlation}$<1% FDR in either men or women were taken forward for determination of $P_{correlation\text{-}sexdiff}$ (Fisher, 1915, 1921; Benjamini and Hochberg, 1995).

3.3.2. Functional enrichment analysis highlights hormones and reproductive traits

A considerable part of all sex-differential aptamer-aptamer correlations seemed to be driven by a relatively small subset of aptamers (Figs 3.8-3.9). Prostate-specific antigen (PSA) alone had evidence of having sex-differential correlations ($P_{correlation\text{-}sexdiff}$<1% FDR) with 1,841 other

aptamers (Fig 3.9) (Fisher, 1915, 1921; Benjamini and Hochberg, 1995). Among the 26 aptamers that were sex-differentially correlated with >5% of other aptamers, I noticed that several seemed to be related to hormonal regulation (LH and FSH) as well as obesity (for example leptin and the growth hormone receptor) (Carmo-Silva and Cavadas, 2017; Matafome and Seiça, 2017). I therefore decided to do a post-hoc functional enrichment analysis using g:GOSt (Raudvere *et al.*, 2019).

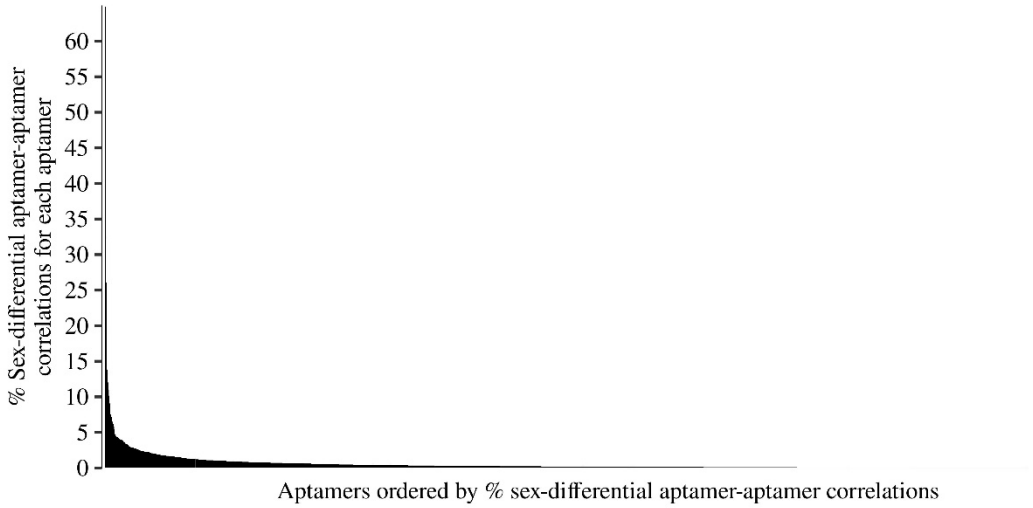


Figure 3.8. Percent sex-differential aptamer-aptamer correlations for each aptamer. Only aptamer pairs that were correlating ($P_{correlation}$<1% FDR) in either men or women were evaluated for sex-differential correlations (Benjamini and Hochberg, 1995).

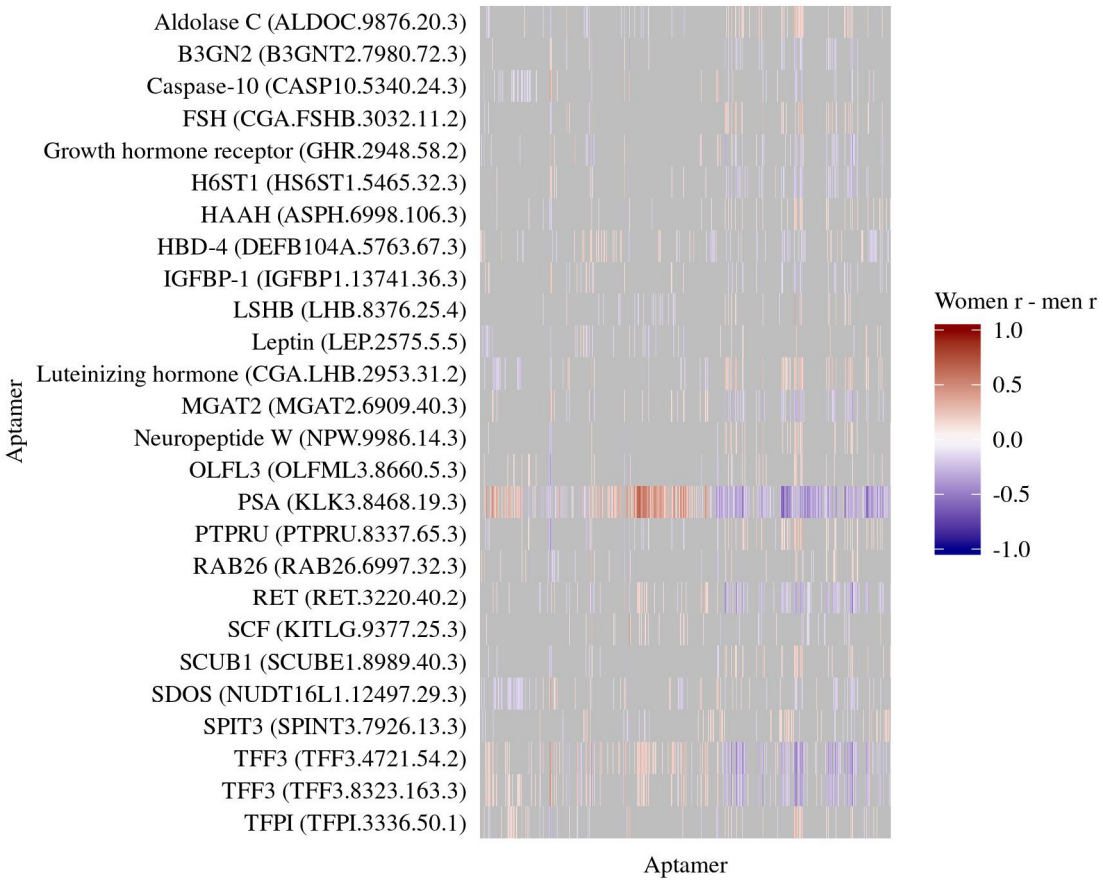


Figure 3.9. The subset of 26 aptamers with evidence of sex-differential aptamer-aptamer correlations with >5% of other investigated aptamers. The Y-axis lists the 26 aptamers with considerable (>5%) sex-differential aptamer-aptamer correlations and the other aptamers in each pair is on the X-axis, unlabelled due to space constraints. Only aptamer pairs that are sex-differentially correlating ($P_{correlation\text{-}sexdiff}$<1% FDR) have the sex difference in the correlation *r* values colour-coded; aptamer pairs without evidence of sex-differential correlation are shown in grey for clarity (Benjamini and Hochberg, 1995).

Using a Bonferroni-corrected multiple-testing threshold for enrichment ($P_{enrichment}$<0.05) in g:GOSt, I identified 27 enriched features (Table 3.4) (Bonferroni, 1936; Raudvere *et al.*, 2019). Of these enriched features, a majority seemed to be related to reproductive and/or hormonal phenotypes, including menstrual cycle aberrations, changed testosterone levels, and the GnRH signalling pathway.

Table 3.4. Enriched features in the functional enrichment analysis of the genes coding for the proteins sex-differentially correlating in >5% of all aptamer pairs.

Term name	Source	Term ID	Bonferroni-adjusted P	Term size	Query size	Intersection size	Precision	Recall
Neuroactive ligand-receptor interaction	KEGG	KEGG:04080	2.9×10^{-4}	55	24	6	0.25	0.11
Reactions specific to the complex N-glycan synthesis pathway	Reactome	REAC:R-HSA-975578	6.9×10^{-4}	5	24	3	0.13	0.60
Glycoprotein hormones	Reactome	REAC:R-HSA-209822	0.002	7	24	3	0.13	0.43
Hormone ligand-binding receptors	Reactome	REAC:R-HSA-375281	0.002	7	24	3	0.13	0.43
Pituitary gland; cells in anterior [≥Medium]		HPA:0370492	0.003	17	24	4	0.17	0.24
Decreased circulating androgen level	Human phenotype ontology	HP:0030349	0.003	13	24	4	0.17	0.31
Decreased serum testosterone level	Human phenotype ontology	HP:0040171	0.003	13	24	4	0.17	0.31
Pituitary gland; cells in anterior [≥Low]		HPA:0370491	0.004	18	24	4	0.17	0.22
Pituitary gland		HPA:0370000	0.004	18	24	4	0.17	0.22
Ovarian steroidogenesis	KEGG	KEGG:04913	0.005	11	24	3	0.13	0.27
Peptide hormone biosynthesis	Reactome	REAC:R-HSA-209952	0.006	9	24	3	0.13	0.33
Abnormal serum testosterone level	Human phenotype ontology	HP:0030087	0.006	15	24	4	0.17	0.27
Peptide hormone metabolism	Reactome	REAC:R-HSA-2980736	0.007	26	24	4	0.17	0.15
Abnormal female reproductive system physiology	Human phenotype ontology	HP:0030012	0.007	57	24	6	0.25	0.11
Mineralocorticoid biosynthesis	Reactome	REAC:R-HSA-193993	0.008	2	24	2	0.08	1.00
Abnormal circulating androgen level	Human phenotype ontology	HP:0030347	0.009	16	24	4	0.17	0.25
Follicle-stimulating hormone complex	Gene ontology, cellular component	GO:0016914	0.010	2	24	2	0.08	1.00
Pituitary gonadotropin complex	Gene ontology, cellular component	GO:0061696	0.010	2	24	2	0.08	1.00
Prader-Willi and Angelman Syndrome	WikiPathways	WP:WP3998	0.011	14	24	3	0.13	0.21

Term name	Source	Term ID	Bonferroni-adjusted P	Term size	Query size	Intersection size	Precision	Recall
Follicle-stimulating hormone activity	Gene ontology, molecular function	GO:0016913	0.012	2	24	2	0.08	1.00
Abnormal circulating hormone level	Human phenotype ontology	HP:0003117	0.015	64	24	6	0.25	0.09
Androgen biosynthesis	Reactome	REAC:R-HSA-193048	0.025	3	24	2	0.08	0.67
GnRH signaling pathway	KEGG	KEGG:04912	0.027	19	24	3	0.13	0.16
hsa-miR-4283	miRTarBase	MIRNA:hsa-miR-4283	0.028	11	24	3	0.13	0.27
N-glycan antennae elongation in the medial/trans-Golgi	Reactome	REAC:R-HSA-975576	0.036	16	24	3	0.13	0.19
Abnormality of the menstrual cycle	Human phenotype ontology	HP:0000140	0.041	46	24	5	0.21	0.11
Abnormality of the ovary	Human phenotype ontology	HP:0000137	0.045	47	24	5	0.21	0.11

Analysis performed using g:OST (Raudvere *et al.*, 2019). Term size, the total number of genes attributed to the term; Query size, the number of genes included in the query available in g:GOSt; Intersection size, the number of genes queried that have been attributed to the term; Precision, Intersection size/query size; Recall, Intersection size/Term size.

3.3.3. Sex-stratified genome-wide association studies of protein levels detect novel, sex-specific effects

I identified 2,295 independent pQTLs (P_{GWAS}<1% FDR, corresponding to $P_{GWAS}\leq 1.1\times10^{-7}$) for 1,495 aptamers in women targeting 1,426 proteins or protein complexes (Benjamini and Hochberg, 1995). For men, 2,426 pQTLs associated with a total of 1,546 aptamers surpassed the multiple testing threshold, targeting a total of 1,462 protein or protein complexes (Benjamini and Hochberg, 1995). Of the pQTLs identified in women, 16% (375 pQTLs) had sex-differential estimates ($P_{sexdiff}$<1% FDR, corresponding to $P_{sexdiff}\leq 2.7\times10^{-4}$) (Appendix Tables 3.1 and 3.2) (Benjamini and Hochberg, 1995; Randall *et al.*, 2013). In men, 15% (354) of the identified pQTLs had sex-differential effect estimates (Benjamini and Hochberg, 1995; Randall *et al.*, 2013).

Of these, three pQTLs associated with ephrin-B3, angiopoietin-4, and Zinc finger protein SNAI2, respectively, had seemingly opposite effect directions on the protein levels in men versus women and would not have been identified in a sex-combined analysis (P_{GWAS}<0.05) (Fig 3.10). Although the sex-specific associations did not surpass P_{GWAS}<1% FDR in both sexes, the opposing effect directions resulted in an apparent null association in the sex-combined analysis.

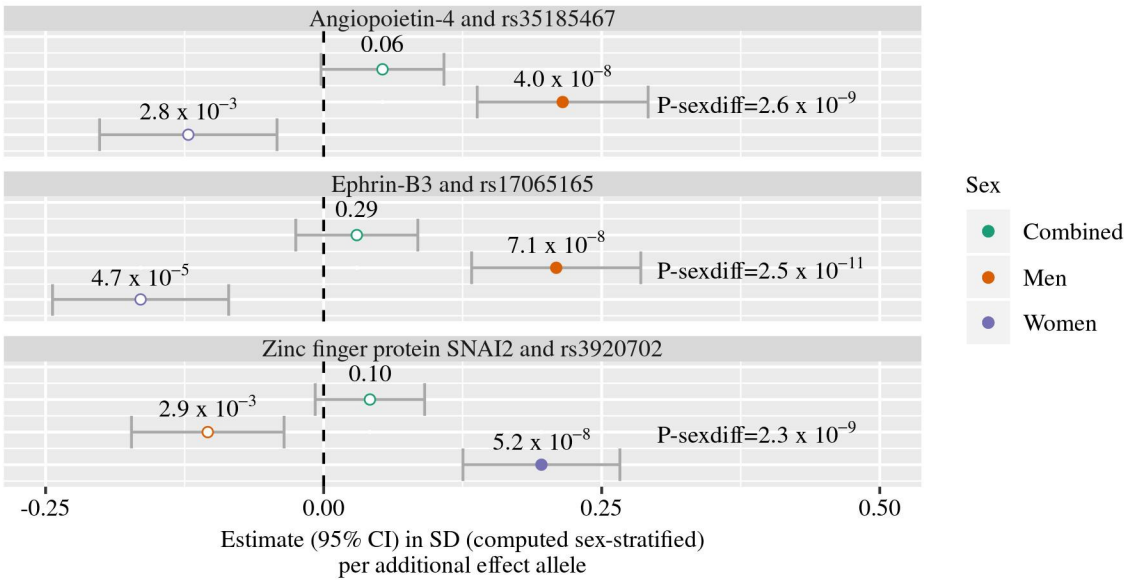


Figure 3.10. pQTLs with a null effect in the sex-combined meta-analysis, seemingly because of opposite effect directions on protein levels in men versus women. The SNPs rs17065165 and rs35185467 are in high LD (r^2=1 in Europeans) and represent the same locus (Machiela and

Chanock, 2015). Filled point estimates indicate P_{GWAS}<1% FDR (Benjamini and Hochberg, 1995). Uncorrected P_{GWAS}-value for each analysis over the point estimate and $P_{sexdiff}$-values shown on the plot (Randall *et al.*, 2013; Winkler *et al.*, 2015).

However, two of the pQTLs, rs17065165 (associated with ephrin-B3) and rs35185467 (associated with angiopoietin-4) are in high LD with each other in Europeans (r^2=1) (Machiela and Chanock, 2015). Thus, the two SNPs represent the same locus with opposite effect directions on both ephrin-B3 and angiopoietin-4, strengthening evidence of a true sexually antagonistic locus. However, the associations between the rs17065165 EA (A) and ephrin-B3 using another aptamer (EFNB3.7785.1.3) did not support the findings (men β -0.030, 95% CI -0.107 to 0.046, P_{GWAS}=0.44, and women β 0.081, 95% CI 0.001 to 0.160, P_{GWAS}=0.05).

3.3.4. Genetic aptamer-aptamer correlation analyses

Of the 2,840 investigated aptamers, only 332 aptamers had a mean $chi^2 \geq 1.02$ in at least one sex and were taken forward for genetic correlation analysis (B. Bulik-Sullivan *et al.*, 2015; B. K. Bulik-Sullivan *et al.*, 2015; Bulik-Sullivan and Finucane, 2017a). However, the LDSC software was only able to estimate genetic correlations for 55,223 aptamer pairs of the total number of 164,838 analyses performed in all sex-strata (B. Bulik-Sullivan *et al.*, 2015). Of these 55,223 pairs, only 11 aptamer-aptamer combinations had a genetic correlation P-value (P_{LDSC})<1% FDR (corresponding to $P_{LDSC} \leq 3.2 \times 10^{-8}$). Of these 11 aptamer pairs, only one pair had genetic correlation estimates available in both sexes (for protein S100-A12 and gamma-aminobutyric acid receptor-associated protein-like 1, measured by S100A12.5852.6.3 and GABARAPL1.12661.44.3, respectively). However, this aptamer pair had highly similar genetic correlation estimates in men (rg=0.99, P_{LDSC}=0.40) and women (rg=0.98, $P_{LDSC}=3.5 \times 10^{-36}$). I therefore decided not to investigate the evidence of sex differences in the genetic correlation analyses any further.

3.3.5. Sex-stratified Mendelian randomisation analyses identify sex-specific effects of proteins on obesity

It was possible to generate instruments in both men and women for 386 aptamers, which were then investigated for associations with BMI, WHR, WHRadjBMI, T2D, and CAD. In total, 43 protein-outcome pairs in men and 48 pairs in women – 84 unique pairs – surpassed the P_{MR}<1% FDR multiple testing threshold (corresponding to $P_{MR} \leq 1.9 \times 10^{-4}$; Appendix Tables 3.3-3.5) (Benjamini and Hochberg, 1995). Several of the results confirmed previous MR findings, such as links between the interleukin-6 receptor and CAD (Swerdlow *et al.*, 2012; Bretherick *et al.*, 2020; Folkersen *et al.*, 2020). Most protein-outcome pairs affected men and women similarly. For example, higher neuroendocrine convertase 1 (PCSK1) levels decreased BMI in both sexes. However, of the 84 unique protein-outcome pairs, 16 combinations (19%) had evidence of having sexually heterogenous effect estimates (P_{het}<1% FDR, corresponding to $P_{het} \leq 0.001$). After excluding protein-outcome combinations that were not supported by sensitivity analyses (see Fig 3.5 for criteria), seven proteins with robust sexually heterogenous effects on obesity remained (Fig 3.11).

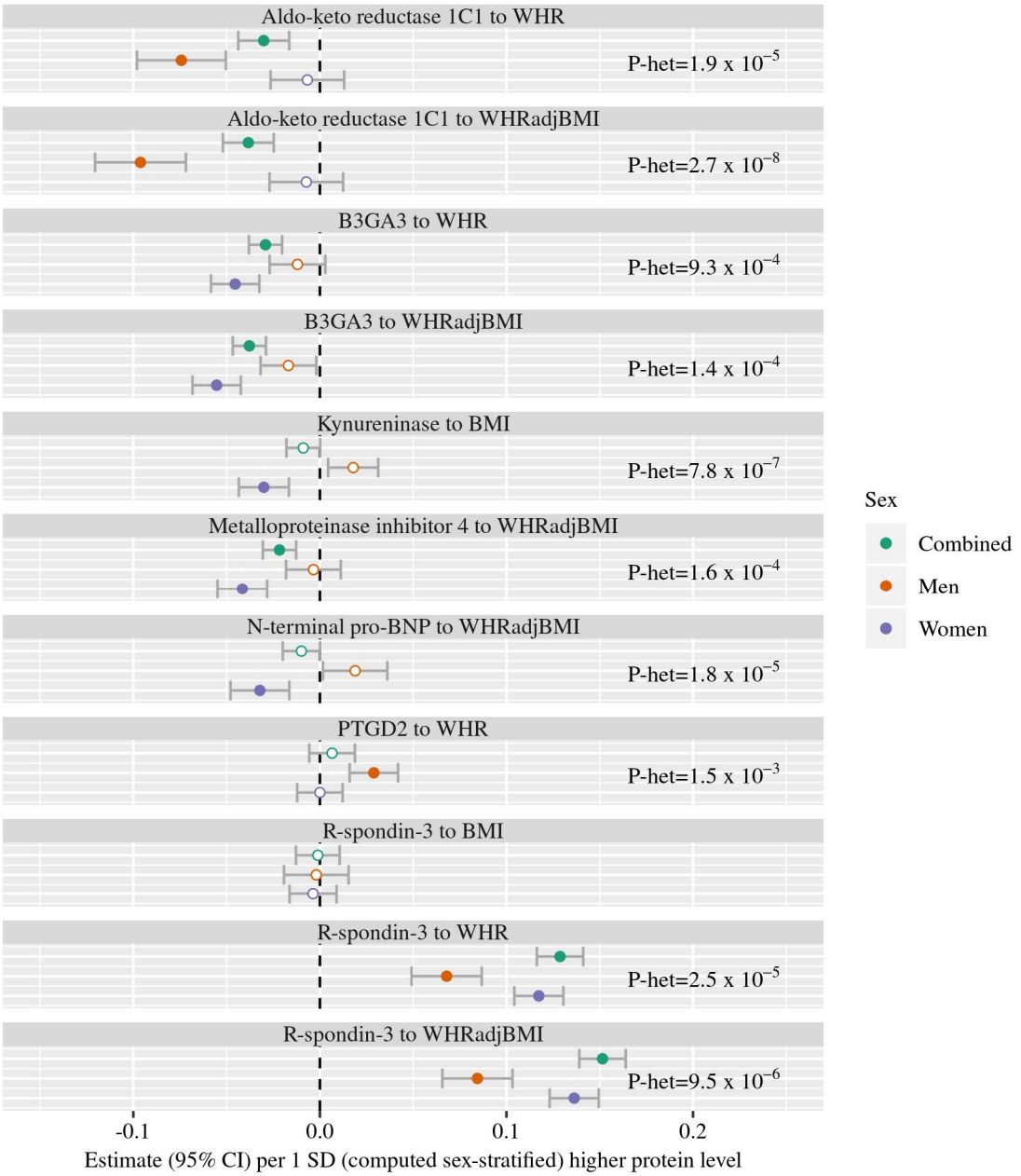


Fig 3.11. Proteins with sexually heterogenous effect estimates on obesity. Only protein-outcome pairs where sexual heterogeneity was supported in the sensitivity analyses are included, see section 3.2.8.2 for criteria, as well as the r-spondin-3 to BMI association for comparison (no P_{het}-value was computed for this analysis since no sex-strata surpassed the P_{MR}<1% FDR criterium). Note that the sex-combined point estimates for r-spondin-3 are not between the sex-stratified point estimates due to the sex-combined analyses being conducted using index instruments identified in the sex-combined analysis, rather than meta-analysis of the sex-stratified MR results. Hence, the genetic variants included in each analysis may differ by sex-strata; for sex-comparisons using the same SNPs as instruments see Appendix Table 3.4. Filled point estimates indicate P_{MR}<1% FDR in that analysis (Benjamini and Hochberg, 1995). Uncorrected P_{het}-values shown on the plot (Cochran, 1954). B3GA3, Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3; N-terminal pro-BNP, N-terminal pro b-type natriuretic peptide; PTGD2, Hematopoietic prostaglandin D synthase.

The relationships between the men and women effect estimates differed for these seven proteins and their effects on obesity. For example, some proteins seemed to have an effect on obesity in a single-sex, such as PTGD2. Other proteins, such as R-spondin-3, seemed to increase the obesity trait in both men and women, but to different degrees. Finally, the results gave suggestive evidence that two proteins – N-terminal pro-BNP and kynureninase – decreased obesity in women, but increased obesity in men, albeit not at P_{MR}<1% FDR in both sexes (Benjamini and Hochberg, 1995).

3.3.6. Sensitivity analyses

3.3.6.1. Reconducting the observational aptamer-aptamer correlations using a different protein transformation approach yields similar results

The observational correlation results using another protein transformation approach, in which the protein levels had first been RINTed before being re-normalised, yielded highly similar results as the main analysis (Appendix Fig 3.1) (Pain, Dudbridge and Ronald, 2018). The subset of aptamers with evidence of sex-differential correlations with >5% of the other aptamers remained in this subset using this other protein transformation approach, with the exception of LHB.8376.25.4 (targeting Lutropin subunit beta in LH). However, the other aptamer targeting the full LH complex (CGA.LHB.2953.31.2) was still sex-differentially correlated with >5% of all other aptamers.

3.3.6.2. Poor agreement between the aptamers

In total, there were 240 unique aptamer pairs where both aptamers targeted the same protein (196 unique proteins). Of these, 142 pairs (59%) in women and 144 pairs (60%) in men were positively correlating $P_{correlation}$<1% FDR using observational data (Benjamini and Hochberg, 1995).

The support for the identified pQTLs using other aptamers targeting the same protein was generally poor. Of the 2,295 pQTLs identified in women, 339 (15%) pQTLs measured proteins that were targeted by two or more aptamers. In men, 384 out of 2,246 pQTLs (16%) measured proteins that were targeted by several aptamers. In these relatively small subsets, only a small

proportion of the pQTLs were supported in the analyses using the other aptamers measuring the same protein (Table 3.5). Only 39% of the pQTLs identified in women and 43% of the pQTLs identified in men were supported in the GWASs of the other available aptamers, i.e., associated with the protein with consistent effect directions at $P_{GWAS}<0.05$. In addition, I found that the proportion of pQTLs with support from other aptamers was even lower among the pQTLs with sex-differential effect estimates ($P_{sexdiff}<1\%$ FDR) (Benjamini and Hochberg, 1995; Randall *et al.*, 2013). Only 30% of the sex-differential pQTLs identified in men, and 27% of the sex-differential pQTLs identified in women, replicated at $P_{GWAS}<0.05$ for all aptamers measuring the same protein.

Table 3.5. The proportion of pQTLs with consistent support from results using all other aptamers targeting the same protein.

pQTL subset	Men	Women
All pQTLs, $P_{GWAS}<0.05$	164 of 384 (43%)	133 of 339 (39%)
Sex-differential pQTLs, $P_{GWAS}<0.05$	19 of 64 (30%)	16 of 59 (27%)
All pQTLs, $P_{GWAS}<1\%$ FDR	90 of 384 (23%)	72 of 339 (21%)
Sex-differential pQTLs, $P_{GWAS}<1\%$ FDR	5 of 64 (8%)	1 of 59 (2%)

Only pQTLs targeting a protein or protein complex quantified by several aptamers are included. Consistent support is denoted as the same EA being associated with higher protein levels in the same sex in the analyses of all other aptamers measuring the same protein at the specified significance threshold, either a nominal threshold of $P_{GWAS}<0.05$ or the multiple testing corrected threshold used in the main analysis of $P_{GWAS}<1\%$ FDR (corresponding to $P_{GWAS}\leq1.1\times10^{-7}$) (Benjamini and Hochberg, 1995).

3.3.6.3. No effect of rs17065165 and rs35185467 on deep venous thrombosis or blood pressure

Neither rs17065165 nor rs35185467, the two pQTLs with seemingly opposite effects on ephrin-B3 and angiopoietin-4 in men versus women, were associated ($P_{GWAS}<0.05$) with deep venous thrombosis, SBP, or DBP in either sex using publicly available summary statistics (Neale lab, 2020).

3.3.6.4. Protein transformation approach can have a substantial impact on results, and sex differences may be affected by age

Of the 729 pQTLs with evidence of having sex-differential effect estimates in men and/or women (some representing the same locus), only 130 pQTLs passed the criteria for having robust evidence of sex-differential effect estimates (see section 3.2.9.4.2. for criteria). Close inspection of the results indicated that protein transformation approach and, consequently, unit of comparison could

have a substantial impact on the results. Whether a change in protein transformation approach affected the results or not seemed primarily dependent on the relative protein distribution in men versus women, which differed between proteins.

One of the most complicated proteins in this regard was LH, measured using the CGA.LHB.2953.31.2 aptamer. Mean LH distribution differed between the two cohorts and between the sexes (Fig 3.12). There was a sex difference in the skew and number of outliers in the data. Furthermore, the distribution in women had a larger variance and was bimodal, with the bimodal peaks seemingly dependent on age or an age-related factor. Consequently, a 1 sex- and cohort-specific SD unit corresponded to a larger amount of log-transformed RFUs in women (small cohort SD=0.82; large cohort SD=0.84) than in men (small cohort SD=0.30; large cohort SD=0.31). Similarly, one SD corresponded to a slightly higher amount of log-transformed RFUs in the larger cohort than the smaller cohort in both sexes.

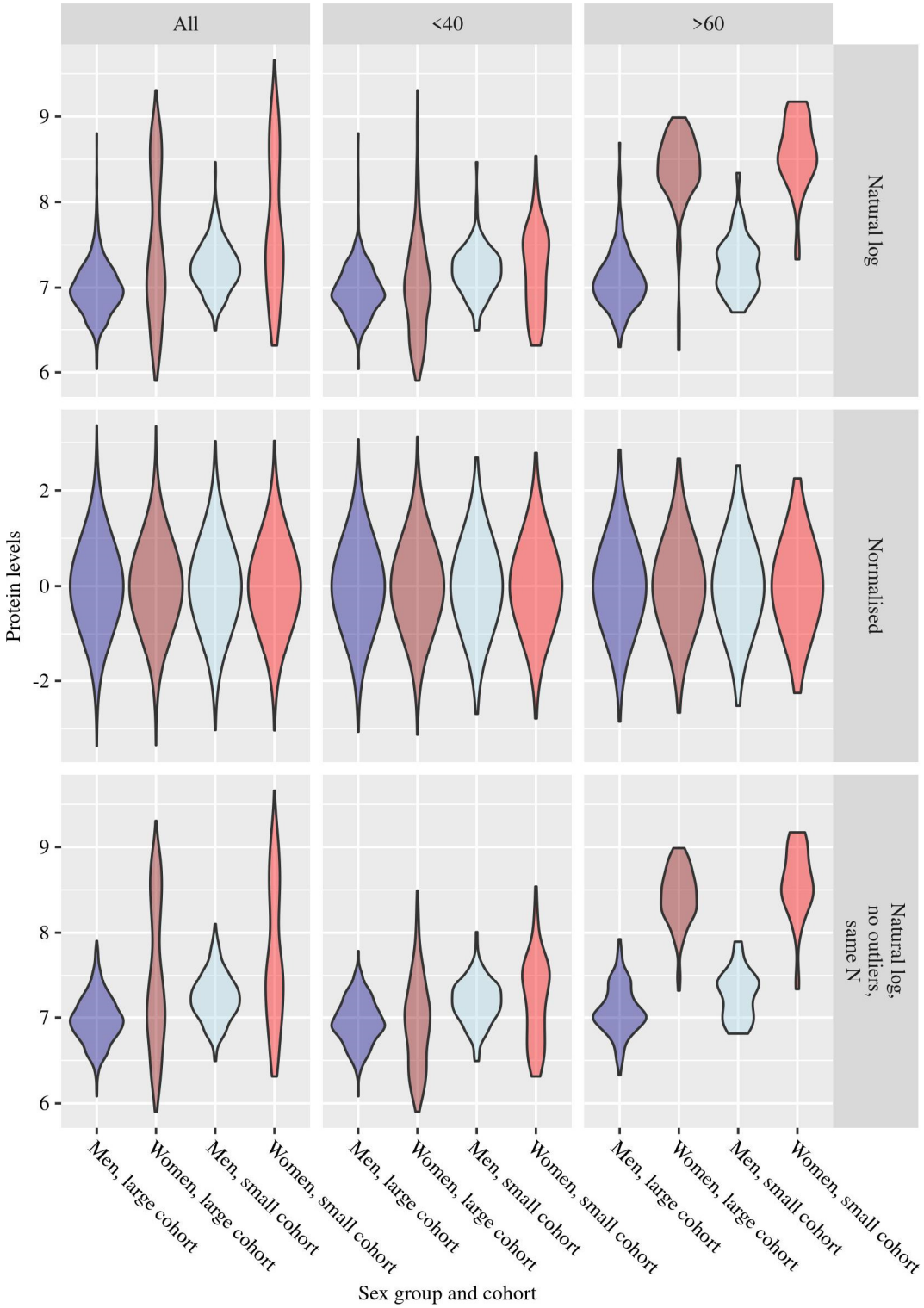


Figure 3.12. Distribution of LH per cohort and sex for the different age-strata, depending on protein transformation approach. For the natural log transformed distributions, the LH values in RFUs were natural log transformed for each sample. For the normalised protein distributions, LH were transformed in the same was as in the main analysis, using RINT of the residuals from a linear regression of the protein and adjusting for age, blood processing time, and 3 PCs, conducted separately in each sex and cohort (Bliss, 1967; Sun *et al.*, 2018). For the natural log transformed distributions without outliers and with the same N, samples with LH values >3 SD (Osborne, 2017c) from the mean (per sex and cohort) were excluded and the sex with the largest sample size randomly subset to contain the same number of samples as the other sex, before the LH values in RFUs were natural log transformed.

These differences in LH distributions led to substantial effects on the apparent sex differences when using different protein transformation procedures. The pQTL rs78537284 G-allele had strong evidence ($P_{sexdiff}=9.7\times10^{-10}$) of being associated with *higher LH levels in men* than in women when using the main protein transformation approach with effect estimates per 1 sex-specific SD unit (Fig 3.13 and Appendix Fig 3.2) (Randall *et al.*, 2013; Winkler *et al.*, 2015). However, when recomputing the effect estimates using the natural log approach (with exclusion of outliers and with the same number of men and women), there was nominal evidence ($P_{sexdiff}=0.046$) of the pQTL being associated with *higher LH levels in women* than in men.

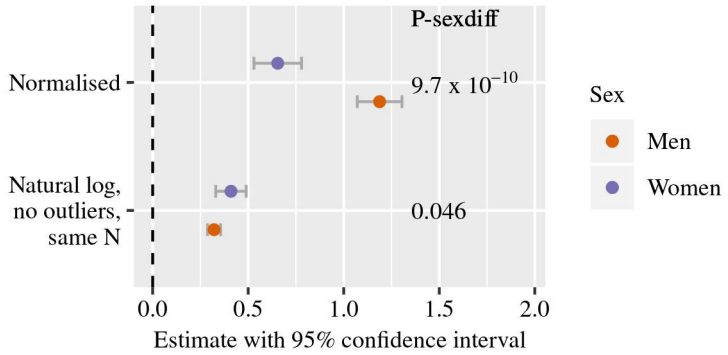


Figure 3.13. Association between rs78537284 and LH levels depending on protein transformation approach and unit of comparison. LH was measured with the aptamer CGA.LHB.2953.31.2 and the LH increase (in sex-specific SD units for the normalised analysis and in log-transformed RFUs for the natural log analysis) per G-allele for the pQTL rs78537284. For the normalised association, LH was transformed in the same way as in the main analysis using RINT of the residuals from a linear regression of the protein, adjusting for age, blood processing time, and 3 PCs, conducted separately in each sex and cohort (Bliss, 1967; Sun *et al.*, 2018). For the natural log transformed association, without outliers and with the same N, samples with LH values >3 SD (Osborne, 2017c) from the mean (per sex and cohort) were excluded and the sex with the largest sample size randomly subset to contain the same number of samples as the other sex, before the LH values in RFUs were natural log transformed. Adjustments for age, blood processing time, and 3 PCs were performed in the genetic association analysis. Uncorrected $P_{sexdiff}$-values shown on the plot (Randall *et al.*, 2013; Winkler *et al.*, 2015).

For other proteins, such as protein FAM19A4, sex differences between men and women seemed to differ between the age-groups (Fig 3.14). However, for these analyses it should be noted that the sample sizes were very low; up to 595 men and 747 women in the <40 years old age-group, and up to 314 men and 170 women in the >60 years old age-group.

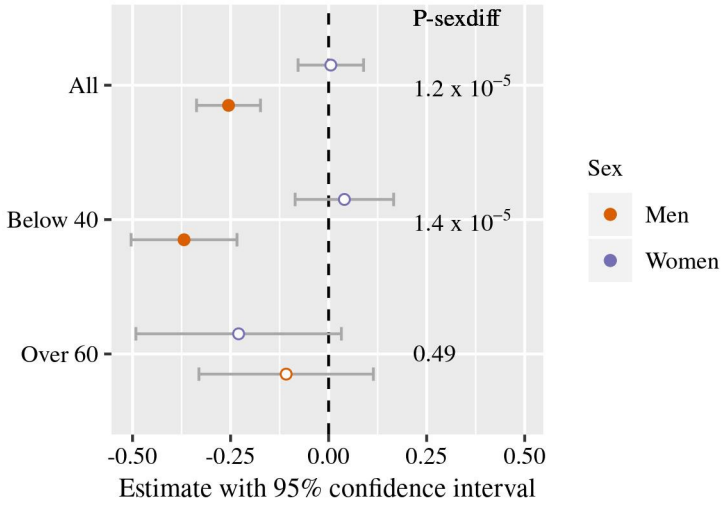


Figure 3.14. Association between rs12500806 and protein FAM19A4 depending on age-group. Protein FAM19A4 was measured with the aptamer FAM19A4.6511.17.3 and the protein FAM19A4 increase (in sex-specific SD units) per C-allele for the pQTL rs12500806. In this analysis, LH was transformed in the same was as in the main analysis using RINT of the residuals from a linear regression of the protein, adjusting for age, blood processing time, and 3 PCs, conducted separately in each sex, cohort, and age-strata (Bliss, 1967; Sun *et al.*, 2018). Filled point estimates indicate P_{GWAS}<1% FDR in that analysis (Benjamini and Hochberg, 1995). Uncorrected $P_{sexdiff}$-values shown on the plot (Randall *et al.*, 2013; Winkler *et al.*, 2015).

In the main sex-stratified GWASs, three pQTLs for angiopoietin-4, ephrin-B3, and Zinc finger protein SNAI2 seemed to have opposite effect directions in men versus women. The sensitivity analyses using the different protein transformation approaches and adjustment protocols largely supported the pQTLs having opposite effect directions in the two sexes (Appendix Figs 3.3-3.5). For all three pQTLs, the men and women point estimates had different effect directions regardless of protein transformation and adjustment protocol in the full age-strata and at least all <40 age-strata analyses.

3.3.6.5. Bidirectional relationships between proteins and cardiometabolic traits and diseases

The bidirectional MR results indicated that fifteen proteins were also affected (P_{MR}<0.05) by obesity and/or cardiometabolic disease (Appendix Table 3.6). Some of these proteins were in the subset of proteins with robust sexual heterogeneity in their effect on obesity, including aldo-keto reductase 1C1, R-spondin-3, and N-terminal pro-BNP (P_{MR} between 0.01-0.04).

3.3.6.6. No evidence of sex-differential participation bias in the INTERVAL data

No SNP was associated with sex (P_{GWAS}<5% FDR) in either cohort or in the meta-analysis of the two cohorts.

The mean chi^2-value was only 0.998 in the GWAS of sex in INTERVAL. As a mean chi^2-value <1.02 is indicative that the data is not suitable for LDSC, I did not move forward with computing genetic heritability estimates of sex (B. K. Bulik-Sullivan *et al.*, 2015; Bulik-Sullivan and Finucane, 2017a).

3.4. Discussion and Conclusion

My main research aim for this chapter was to identify potential sex-specific protein biomarkers and drug targets for cardiometabolic traits and diseases. To achieve this aim, I had two primary research objectives; to identify sex-specific pQTLs and to investigate sex-specific effects of proteins on obesity and leading cardiometabolic diseases using MR. Sex-stratified GWASs of protein levels identified hundreds of pQTLs with sex-differential effects on protein levels (Randall *et al.*, 2013). Three pQTLs even seemed to increase protein levels in one sex, but decrease protein levels in the other sex, resulting in a null association in the sex-combined meta-analysis. I also identified seven proteins with robust sexually heterogenous effects on obesity (Cochran, 1954), of which two proteins likely increase obesity in one sex but decrease obesity in the other sex. These results suggest that these seven proteins may be useful as biomarkers and/or drug targets in obesity, but that their clinical utility may differ between the sexes. In addition, my results emphasise the need for sex-stratified analyses in genetic association studies in order to identify and characterise associations that differ between the sexes.

3.4.1. Protein-protein correlations in the human plasma proteome

As many proteins interact with each other and share functional features (Alberts *et al.*, 2014; Szklarczyk *et al.*, 2019), I first investigated observational correlations between aptamer pairs in

each sex (Fisher, 1915, 1921; Osborne, 2017c). It should be noted that the small subset of aptamer pairs where both aptamers targeted the same protein would be expected to have a high degree of correlation. Yet, the results indicated wide-spread correlations between the majority of all aptamer pairs, and with few differences between the sexes. Given this large degree of correlation between the aptamers and the sexes, I used an FDR approach to correct for multiple testing in this chapter (Benjamini and Hochberg, 1995).

I found that a small subset of proteins was sex-differentially correlated in a relatively large proportion of all correlation analyses. There are several possible explanations for this finding. Firstly, these proteins may be sex-differentially regulated by, for example, hormones or transcription factors which otherwise have similar effects on a large part of the remaining proteome. The functional enrichment analysis of these proteins supported this theory, as the results suggested that the proteins were enriched for hormonal regulation and reproductive phenotypes. Furthermore, a previous study has suggested that many sex-differentially expressed genes are regulated by hormone-related transcription factors (Oliva *et al.*, 2020). Another possible explanation is that these proteins are sex-differentially affected by a common, unadjusted confounder that otherwise has similar effects on the proteome in men and women. However, I could not identify any obvious common confounders with different effects on the identified proteins in men versus women in the literature. Thirdly, I hypothesised that the pattern might be caused by bias resulting from the protein transformation approach. In brief, transforming data using a standard RINT of the residuals may re-introduce associations with adjusted confounders if the data is skewed or has outliers (Pain, Dudbridge and Ronald, 2018). I therefore conducted a sensitivity analysis using a different protein transformation approach that should be more robust to such bias, but the results were highly similar as to the main analysis (Pain, Dudbridge and Ronald, 2018). Finally, it is possible that the results were affected by non-specific aptamer binding. For example, if an aptamer can bind to two different proteins and the relative levels of these proteins differ between men and women, the aptamer is effectively measuring different proteins in the two sexes. We would then expect wide-spread sex differences in the correlations for that aptamer. In

summary, there are many possible explanations as to why some proteins exhibited wide-spread sex-differential correlations with other proteins, and it is likely that several of these explanations contribute to the results.

3.4.2. Sex-stratified genome-wide association studies of protein levels

3.4.2.1. Sex differences in the effect estimates of protein quantitative trait loci are common

Of the pQTLs identified in each sex, approximately 15% had evidence of having sex-differential associations with the protein. While there has not been any previous large-scale studies investigating the evidence of sex-specific pQTLs, a study by Oliva *et al.* (2020) and another study by Porcu *et al.* (2020) assessed the evidence for eQTLs having different effects in men versus women. Oliva *et al.* (2020) identified 369 sex-differential eQTLs for 366 genes across 44 tissue types (FDR<25%) (Benjamini and Hochberg, 1995). In the study by Porcu *et al.* (2020), which was conducted using expression data from a single tissue type (whole blood), 18 genes with sex-differential eQTLs were identified (FDR<5%) (Benjamini and Hochberg, 1995). However, contrary to my approach where I investigate sex differences in pQTLs that are associated with the proteins P_{GWAS}<1% FDR in *each sex*, both Oliva *et al.* (2020) and Porcu *et al.* (2020) only investigated sex differences in the eQTLs that were associated with gene expression in a *sex-combined* analysis. The downside of such an approach is that genetic variants that have opposite effects on the trait in men versus women, or an effect in a single sex, may not surpass the multiple testing threshold in a sex-combined analysis. In my results, I find evidence that three trans-pQTLs, with seemingly opposite effect directions in men versus women, even had a null association in the sex-combined meta-analysis. This may explain why I find evidence of a much larger proportion of quantitative trait loci having sex-differential effects, but other possible explanations are differences between expression and protein data or between tissue types (Oliva *et al.*, 2020).

3.4.2.2. Genetic variants may be associated with higher protein levels in one sex, but lower protein levels in the other sex

Of the three trans-pQTLs with a null association in the sex-combined analysis, the pQTL with seemingly opposite effects on ephrin-B3 levels in the two sexes was not supported using another aptamer targeting ephrin-B3. However, knock-out of the ephrin-B3 gene in mice seem to increase blood pressure in female mice only (Wang *et al.*, 2016). In addition, the locus also had similar sex-differential effects on angiopoietin-4, suggesting that it is the contradicting ephrin-B3 aptamer that is of poor quality. As both ephrin-B3 and angiopoietin-4 are implicated in angiogenesis and ephrin-B3 has been linked to blood pressure (Yancopoulos *et al.*, 2000; Olsen *et al.*, 2006; Pfaff, Fiedler and Augustin, 2006; Wang *et al.*, 2016), I investigated if there was evidence of the locus having opposite effects on blood pressure or deep venous thrombosis. However, there was no evidence of the locus having any effects on these traits in humans, wherefore these analyses did not bring any greater clarity in the matter. Still, the results suggest that pQTLs may increase protein levels in one sex and decrease protein levels in the other sex, and as a result potentially be missed in a sex-combined analysis. Thus, my results emphasise the importance of sex-stratified GWASs to identify SNPs with sex-differential associations.

3.4.3. Mendelian randomisation analyses

3.4.3.1. Sex-specific effects of proteins on obesity

Reassuringly, some of the MR results confirmed previous studies. For example, the interleukin-6 receptor has previously been linked to risk of CAD in MR studies (Swerdlow *et al.*, 2012; Bretherick *et al.*, 2020; Folkersen *et al.*, 2020), similar to my results. However, the results also identified novel, sex-specific associations and highlighted seven proteins with robust sexually heterogenous effects on obesity, including kynureninase, N-terminal pro-BNP, r-spondin 3, aldo-keto reductase family 1C1, PTGD2, metalloproteinase inhibitor 4, and B3GA3.

3.4.3.1.1. Kynureninase

My results indicated that higher kynureninase levels may decrease obesity in women, but increase obesity in men. Kynureninase is a key enzyme in the kynurenine pathway, which metabolises tryptophan into kynurenine metabolites (Gál and Sherman, 1980; Oxenkrug, 2010). The metabolites can regulate brain receptors, such as N-methyl-D-aspartate (NMDA)-receptors, which in turn have been implicated in appetite regulation (Stone and Perkins, 1981; Perkins and Stone, 1982; Deng *et al.*, 2019). Furthermore, kynurenine pathway activity may indirectly affect serotonin levels, as serotonin is also made from tryptophan (Gál and Sherman, 1980; Oxenkrug, 2010). Serotonin, in turn, is implicated in regulation of both appetite and body temperature (Halford and Blundell, 2000; Crane *et al.*, 2015). The kynurenine pathway has also been linked to cardiometabolic disease in observational studies (Pedersen *et al.*, 2015; Zuo *et al.*, 2016; Song *et al.*, 2017). Although I did not find any evidence of secondary effects of the protein on T2D or CAD, that may be reflective of the relatively small sample size rather than a true null effect. Thus, there are several plausible causal links between kynureninase and obesity, although it is unclear why higher kynureninase levels would have opposite effects on obesity in men versus women.

Previous studies have suggested that the pathway likely can be modulated using both drugs and lifestyle changes (Stone and Darlington, 2002; Agudelo *et al.*, 2014). Thus, the kynurenine pathway holds promise as a potential target for medical intervention to decrease obesity, and potentially cardiometabolic disease, but future studies should evaluate its efficacy and safety separately by sex.

3.4.3.1.2. N-terminal pro-BNP

N-terminal pro-BNP is an inactive peptide that is released together with BNP in response to increased stretching of the walls of the heart (de Bold *et al.*, 1981; Sudoh *et al.*, 1988; Madamanchi *et al.*, 2014). Therefore, it should be noted that it is highly probable that my N-terminal pro-BNP instrument is actually capturing the effects of changes in BNP. Both N-terminal pro-BNP and BNP – commonly referred to as natriuretic peptides – are already used as a biomarkers for heart failure

in both men and women (Madamanchi *et al.*, 2014). In my study, I found N-terminal pro-BNP to be associated with decreased WHRadjBMI in women and nominal evidence of it increasing WHRadjBMI in men. However, I also found nominal evidence in favour of a bidirectional relationship between the protein and abdominal obesity in women.

Observational studies have found N-terminal pro-BNP and BNP levels to be inversely associated with obesity in both sexes (Wang *et al.*, 2004; Das *et al.*, 2005). Although the relationship between natriuretic peptides and obesity is unclear, one proposed mechanism is higher clearance of the peptides in obese individuals because of higher glomerular filtration rates (Brøchner-Mortensen, Rickers and Balslev, 1980; Tsutamoto *et al.*, 2006; Madamanchi *et al.*, 2014). However, there is previous evidence of a bidirectional relationship between natriuretic peptides and obesity. A study in human subcutaneous adipocytes found BNP to increase lipolysis (Sengenès *et al.*, 2000), and a study in mice suggested that the peptides increase fat oxidation and thereby decrease obesity (Miyashita *et al.*, 2009). Yet, it is unclear if and how these potential pathophysiological mechanisms differ between men and women, and why the results indicated opposite effects of the protein on abdominal obesity between the sexes.

3.4.3.1.3. R-spondin 3

R-spondin 3 was found to increase abdominal obesity in both sexes, but to a larger degree in women. Genetic variants close to the gene have previously been associated with WHRadjBMI, with larger effect estimates in women and sex-differential expression of the gene in adipose tissue (Loh *et al.*, 2020). The protein likely affects body fat distribution by regulating adipogenesis, apoptosis susceptibility, and adipose progenitor cell proliferation (Loh *et al.*, 2020). In addition, studies in male rats have indicated that both R-spondin 3 and another member of the r-spondin family, R-spondin 1, may regulate food intake (Li *et al.*, 2014). Thus, my results highlight that r-spondin 3 and its affiliated pathways may be of interest as potential biomarkers and/or drug targets in both men and women.

3.4.3.1.4. Aldo-keto reductase family 1C1 and hematopoietic prostaglandin D synthase affect abdominal obesity in men

Higher levels of aldo-keto reductase family 1C1 decreased abdominal obesity primarily in men, with no clear effect in women. The enzyme converts the sex hormone progesterone to less active forms, and a previous observational study in men found progesterone to be inversely correlated with BMI and waist circumference (Blanchette *et al.*, 2006; El-Kabbani, Dhagat and Hara, 2011; Zeng *et al.*, 2019). However, a larger observational study found no association between progesterone and obesity in neither sex (He *et al.*, 2018). Hence, more studies are needed to determine the role of aldo-keto reductase family 1C1 in obesity and how any sex-specific effects are mediated.

I also found evidence of PTGD2 increasing WHR in men only. PTGD2 synthesises prostaglandin D2, in turn involved in for example platelet aggregation, smooth muscle relaxation and inflammation (Kanaoka and Urade, 2003). Expression of PTGD2 in adipose tissue is increased in obese humans and inhibition of the protein in male mice lead to decreased food intake (Hernandez-Carretero *et al.*, 2018). In addition, mice overexpressing PTGD2 gain more weight (Fujitani *et al.*, 2010). However, it is unclear how PTGD2 would affect abdominal obesity differently in men versus women.

3.4.3.1.5. Metalloproteinase inhibitor 4 and galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3 affect abdominal obesity in women

Two other proteins, metalloproteinase inhibitor 4 and B3GA3, primarily seemed to affect abdominal obesity in women, with no association detected in men. There is evidence linking metalloproteinase inhibitor 4 to adipogenesis (Mejia-Cristobal *et al.*, 2015). A study by Sakamuri *et al.* (2017) found that knock-out mice of the gene had a higher food intake but a lower body fat gain compared to wild-type. For B3GA3, it is unclear if and how the protein affects abdominal obesity; rare mutations in the gene do not seem to substantially affect obesity or adipogenic traits (Ritelli *et al.*, 2019).

3.4.3.2. Strengths and limitations specific to the Mendelian randomisation analyses

There are several limitations inherent to MR that may affect my results. Firstly, the genetic variants must only be associated with the risk factor, and not with any potential confounders (Greenland, 2000; Martens et al., 2006; Burgess et al., 2015). In my subset of "high-confidence" protein-outcome MR analyses, I excluded all analyses where any instrument was associated with any other protein than that under investigation. Yet, it is impossible to verify that the instruments are not associated with unmeasured confounders.

Secondly, I created instruments based on SNPs that surpassed a 1% FDR threshold (Benjamini and Hochberg, 1995), and many proteins had a single-SNP instrument. As a result, it is possible that some of my results suffer from weak instrument bias (Staiger and Stock, 1997; Burgess, Davey Smith, *et al.*, 2020). Yet, weak instrument bias will be in the direction of the null in a non-overlapping two sample MR analysis, wherefore this should primarily result in false negative results (Staiger and Stock, 1997; Burgess, Davey Smith, *et al.*, 2020). Still, sex-differential weak instrument bias could potentially attenuate the MR estimate more in one sex than the other, which could lead to false positive sexual heterogeneity between the sex-specific effect estimates.

3.4.4. Limitations inherent to the data

3.4.4.1. Sex-differential effect estimates may be dependent on unit of comparison and age-related factors

For some proteins, such as LH, the distribution varied between the cohorts, age-groups, and between the sexes. Protein levels were therefore normalised separately by sex and cohort, and hence sex-comparisons were made between effect estimates in sex-specific SD units. If protein distributions differ between the sexes, sex differences might therefore be different if comparing effect estimates in absolute protein units, as exemplified with the LH-associated locus. There may also be sex-differential bias arising from, for example, skew or outliers, confounders, or regression dilution bias (Osborne, 2017a; Pain, Dudbridge and Ronald, 2018).

Yet, the problem is not simply rectified by using another protein transformation approach. I evaluated if the sex-differential pQTLs were still sex-differential using variations of several protein transformation approaches and adjustment protocols. However, it should be noted that many of these protein transformation procedures were likely unsuitable for many proteins because of violations of the linear regression assumptions or other issues (Osborne, 2017a). For example, the SomaLogic data had already been normalised for technical variations (SomaLogic, 2017; Sun *et al.*, 2018). Hence, 1 RFU in one cohort does not necessarily equal the exact same amount of protein in the other cohort. Therefore, it could be problematic to meta-analyse effect estimates in, for example, log-transformed RFUs between the two cohorts, as cohort-differences could be due to either true differences in protein distributions or normalisation artefacts. Further complicating such an approach is that the numbers of men and women samples in each cohort were slightly different, which potentially could lead to false positive sex differences in the case of cohort-differences in protein levels. Although I attempted to overcome this particular issue by subsetting the sample to have the same number of men and women in each analysis, my results could still be biased due to, for example, sex-specific violations of the linear regression assumptions caused by skew in the data (Osborne, 2017a).

Another approach that I evaluated was to normalise proteins sex-jointly in each cohort, as suggested by Pirastu *et al.* (2020). Yet, as discussed in Chapter 2, such an approach could result in a bimodal distribution in the sex with the larger protein variance. In addition, for proteins with substantial sex differences in protein mean and variance, it is questionable if this results in a fair comparison between the sexes.

I also performed the analyses in a <40 and >60 years old age-group. The rationale behind these sensitivity analyses was that sex differences may be affected by menopausal status in women (Rannevik *et al.*, 1995). As I did not have access to data on menopausal status, I decided to stratify analyses by age. Indeed, I found some evidence in favour of sex differences being dependent on

age or an age-related factor, such as menopause. This suggests that my results could be affected by the age-distribution in my sample, further complicated by the fact that the age distributions were slightly different between men and women. In addition, it highlights the potential of sex-differential noise or model fits affecting my results. For example, if menopause greatly affects protein levels in women, this could cause sex-differential noise and model fits that in turn biases the sex-comparisons.

The obesity outcomes in the MRs were in sex-specific SD units. Because of that and given the scope of the project, I decided to only perform the MRs in sex-specific SD units. Yet, choice of phenotype transformation approach would be expected to have a similar impact on the MRs, given that the MR estimates are based on the GWAS summary statistics (Burgess and Thompson, 2015; Burgess *et al.*, 2015).

3.4.4.2. Poor aptamer agreement raises concerns about aptamer specificity

Given the discrepancy between the ephrin-B3 aptamers, I decided to investigate if the results were consistent between different aptamers measuring the same protein target. These sensitivity analyses raised concerns about my results. The proportion of correlating aptamer pairs for pairs targeting the same protein was around 60% in both sexes – approximately the same as the proportion of aptamer pairs correlating overall. I then investigated if the index pQTLs replicated using data from the other aptamers targeting the same protein, if available. These results indicated poor agreement between aptamers for the pQTLs, with the replication rates ranging between 2-43% depending on pQTL subset and replication criteria.

In the original publication of the proteomics data from INTERVAL, Sun *et al* (2018) had evaluated a small subset of the index pQTLs (163 out of 1,927 pQTLs, measuring 116 different proteins) using a different protein quantification method. In this analysis, 65% of pQTLs replicated, which is higher than my replication rates, in spite of my expanded exclusion of all aptamers with any evidence of non-specific binding. It is possible that the discrepancy could be caused by my lower

P_{GWAS}-threshold to denote significance in the GWASs, or potentially by chance in what aptamers were included in the analyses (Sun *et al.*, 2018). Regardless, the poor consistency between aptamers implies that a considerable proportion of results using the SomaLogic aptamer data may be false positives and negatives due to poor aptamer quality. Yet, as a proportion of the pQTL results did indeed replicate between aptamers, some of the results are likely to be robust. The difficulty is in identifying which aptamers are of good quality, since the vast majority of the measured proteins were only targeted by a single aptamer. Thus, my results indicate that results based on the SomaLogic assay should be replicated using another protein quantification method before final conclusions are made.

3.4.4.3. Coding variants may change aptamer affinity

The MR results indicated that higher PCSK1 levels decreased BMI in both men and women. This is in keeping with previous studies, as for example rare non-synonymous coding variants in *PCSK1* may cause monogenic rare forms of obesity (Jackson *et al.*, 1997; Creemers *et al.*, 2012). Yet, this example also illustrates that the results may be affected by changed aptamer binding, as one of the PCSK1 instruments (rs6235) affects the coding sequence of the protein (Sherry *et al.*, 2001; Ward and Kellis, 2012; Machiela and Chanock, 2015). If an index SNP (or a proxy) is affecting the coding sequence of a protein, that could affect the aptamer affinity to the protein and lead to false positive pQTLs (Abhishek and Manuel, 2018; Sun *et al.*, 2018). Yet, it is possible for SNPs to affect both coding sequence and expression levels of a protein, as exemplified by a study on genetic variants in the *PAM* gene by collaborators and I (Thomsen *et al.*, 2018). Therefore, it is difficult to know if an instrument is capturing changed protein levels and/or changed aptamer binding without functional follow-up studies of the specific aptamer and pQTL.

3.4.5. Sensitivity analyses

3.4.5.1. No evidence of sex-differential participation bias in the genome-wide association study

A study by Pirastu *et al.* (2020) suggested that sex-differential participation bias and sex-chromosome homology can bias comparisons between sexes in genetic epidemiology. I therefore conducted a GWAS of sex in the INTERVAL dataset, similar to Pirastu *et al.* (2020). In theory, if a trait is sex-differentially associated with participation in a study, genetic variants associated with that trait would seem associated with sex in a GWAS of sex (Pirastu *et al.*, 2020). However, no SNP had any evidence of being associated with sex in the INTERVAL dataset. Still, the sample size was small, and I therefore had low power to detect sex-differential participation bias.

3.4.5.2. Data were generally unsuitable for genetic correlation analyses and for computing heritability

I attempted to use LDSC to compute both genetic correlations between aptamer pairs and to investigate potential sex-differential participation bias by computing heritability estimates of sex (B. Bulik-Sullivan *et al.*, 2015; B. K. Bulik-Sullivan *et al.*, 2015; Pirastu *et al.*, 2020). As I only had access to imputed genetic data, and since the protein levels had been measured separately for the two cohorts, I had decided to use LDSC since it only requires summary statistics (B. Bulik-Sullivan *et al.*, 2015; B. K. Bulik-Sullivan *et al.*, 2015). However, there were very few proteins that were suitable for use with LDSC. It is recommended that traits should have a mean chi^2-value ≥1.02, which is indicative of a polygenic signal for the trait (Bulik-Sullivan and Finucane, 2017a). In my study, very few protein GWASs surpassed this threshold in both sexes, nor did the GWAS of sex (B. Bulik-Sullivan *et al.*, 2015; Bulik-Sullivan and Finucane, 2017a). In addition, a minimum sample size of 5,000 is recommended to have enough power for LDSC (B. Bulik-Sullivan *et al.*, 2015; Bulik-Sullivan and Finucane, 2017a). Thus, the lack of results from the analyses using LDSC is not surprising.

3.4.6. Conclusion

In this chapter, I highlight genetic variants that are sex-differentially associated with protein levels. I identify several proteins that have sexually heterogenous effects on obesity, including two with seemingly opposite effects on obesity in men versus women. Several of these proteins hold promise as potentially useful drug targets for obesity, although secondary protective effects on cardiometabolic disease are unclear. However, sensitivity analyses highlighted potential issues concerning aptamer specificity, and my results need to be replicated in another study using a different protein quantification method before final conclusions are made. The results nonetheless emphasise the importance of sex-stratified analyses in genetic association and MR analyses to identify sex-specific effects.

Chapter 4. Using colocalisation to identify potential mediators in obesity-related female reproductive diseases

4.1. Introduction

4.1.1. The role of obesity in female reproductive diseases

4.1.1.1. Obesity and the risk of polycystic ovary syndrome

Women with PCOS are more likely to be obese than healthy women of a similar age (Glueck *et al.*, 2005). Studies have suggested that obesity and higher amount of visceral fat increase insulin resistance, and thereby worsen the metabolic and endocrine disruptions in women with PCOS (Lord *et al.*, 2006; Lim *et al.*, 2013). In support of this theory, lifestyle changes and weight loss seem to improve ovulatory function in affected women, possibly through recovered insulin sensitivity (Nybacka *et al.*, 2011).

In their 2012 consensus, the PCOS Consensus Workshop Group established that obesity is associated with the disease – but concluded that *"…its causal role in this condition has yet to be determined"* (Fauser *et al.*, 2012). Since then, MR studies have provided evidence of a causal effect of BMI on risk of PCOS (Day *et al.*, 2018; Brower *et al.*, 2019). The study by Brower *et al.* (2019) also investigated if having PCOS increases BMI, but did not find any evidence in favour of a bidirectional relationship. Still, to my knowledge, no previous MR studies have explored if other obesity-related traits – such as abdominal obesity – affect the disease risk.

4.1.1.2. Limited knowledge of the effect of obesity on miscarriage

Meta-analyses of observational studies have identified an association between obesity and both sporadic and recurrent miscarriage (Boots and Stephenson, 2011; Cavalcante *et al.*, 2019).

However, an MR analysis performed by collaborators did not find any evidence of a causal role of BMI in risk of sporadic miscarriage (Laisk *et al.*, 2020). In addition, it is currently unclear if other obesity-related traits – such as abdominal obesity – have an effect on risk of miscarriage.

4.1.2. The genetics of polycystic ovary syndrome and miscarriage

4.1.2.1. The mediating genes for polycystic ovary syndrome are largely unknown

Polycystic ovary syndrome has a strong genetic component; a twin-based study estimated that hereditary factors explain two thirds of the disease variance (Vink *et al.*, 2006). GWASs to date have primarily been conducted in women of European or Han Chinese ancestry, and a total of 19 risk loci have been identified (Chen *et al.*, 2011; Shi *et al.*, 2012; Day *et al.*, 2015, 2018; Hayes *et al.*, 2015). Some of these risk loci are located close to genes with a plausible role in PCOS pathophysiology, such as the genes coding for the insulin receptor (*INSR*), the FSH-receptor (*FSHR)*, and the LH-receptor (*LHCGR*) (Chen *et al.*, 2011; Shi *et al.*, 2012). However, as exemplified by genetic variants in the *FTO* locus and their associations with obesity in section 1.4.3.1, the closest gene may not always be the mediating gene – even when the closest gene is a likely candidate (Gerken *et al.*, 2007; Fischer *et al.*, 2009; Church *et al.*, 2010; Smemo *et al.*, 2014; Claussnitzer *et al.*, 2015; Barroso and McCarthy, 2019). Still, for most risk loci there are several plausible effector genes, and for the majority of PCOS-associated risk loci the mediating gene remains to be identified and/or confirmed (Day *et al.*, 2018).

4.1.2.2. The lack of large genetic association studies of recurrent miscarriage

Family studies have indicated that female relatives of women with recurrent miscarriage have a higher risk of miscarriage themselves, suggesting – but not concluding – a genetic component in the disease (Christiansen *et al.*, 1990; Kolte *et al.*, 2011). However, studies of the genetics of recurrent miscarriage have been hampered by small sample sizes, inter-study differences in disease criteria, and/or contradictory results between different studies (Christiansen *et al.*, 1990; Wang *et al.*, 2010; Kolte *et al.*, 2011; Pereza *et al.*, 2017; Quintero-Ronderos *et al.*, 2017; Laisk *et al.*, 2020). Prior to the GWAS conducted by my collaborators (Laisk *et al.*, 2020), only a few genome-

wide or whole-exome analyses on recurrent miscarriage had been performed (Wang *et al.*, 2010; Kolte *et al.*, 2011; Quintero-Ronderos *et al.*, 2017). These previous genome-wide and whole-exome analyses have identified different risk loci, likely explained by the sample sizes being below 50 cases or families in each study (Wang *et al.*, 2010; Kolte *et al.*, 2011; Quintero-Ronderos *et al.*, 2017). Thus, even though the GWAS performed by Laisk *et al.* (2020) only included up to 750 cases for recurrent miscarriage, it is considerably larger than previous GWASs of recurrent miscarriage (Wang *et al.*, 2010; Kolte *et al.*, 2011).

4.1.3. Female reproductive diseases – common and with far-reaching consequences, but often overlooked

Female reproductive diseases, such as PCOS and miscarriage, are common (Bozdag *et al.*, 2016; Cohain, Buxbaum and Mankuta, 2017). Both diseases have been linked to severe cardiometabolic consequences, such as T2D for women with PCOS (Moran *et al.*, 2010) and increased risk of cardiovascular disease for women who have experienced miscarriage (Maino *et al.*, 2016). In addition, these diseases have a substantial impact on women's mental wellbeing (Cumming *et al.*, 2007; Weng *et al.*, 2018; Damone *et al.*, 2019). Yet, there is an under-diagnosis and under-treatment of PCOS and its consequences (Broder-Fingert *et al.*, 2009; Sivayoganathan *et al.*, 2011; Gibson-Helm *et al.*, 2017; Wolf *et al.*, 2018). In addition, research into these diseases have been thwarted by small sample sizes and/or poor methodology (Aleman *et al.*, 2005; Broder-Fingert *et al.*, 2009; Wang *et al.*, 2010; Kolte *et al.*, 2011; Sivayoganathan *et al.*, 2011; Chen, Fu and Huang, 2016; Qiao *et al.*, 2016; Pereza *et al.*, 2017; Wahabi *et al.*, 2018; Lim *et al.*, 2019). If obesity is a causal risk factor for these diseases (Glueck *et al.*, 2005; Boots and Stephenson, 2011; Day *et al.*, 2018; Brower *et al.*, 2019; Cavalcante *et al.*, 2019), their prevalence may increase if the obesity rates continue to rise (NCD Risk Factor Collaboration (NCD-RisC), 2016b; GBD 2015 Obesity Collaborators *et al.*, 2017). It is thus crucial that we understand the role of obesity in these diseases and that we identify any disease-mediating genes and proteins that may be used as biomarkers or targeted by drugs to ameliorate the consequences of these diseases.

4.1.4. Aims

My main research aim in this chapter was to establish the effect of obesity on two obesity-associated female reproductive diseases and identify potential disease-mediating genes (Fig 4.1). To explore this question, I had two research objectives:

I. To investigate the causal effect of obesity on PCOS and miscarriage using MR.

II. To perform colocalisation analysis between these female reproductive diseases and genes/proteins.

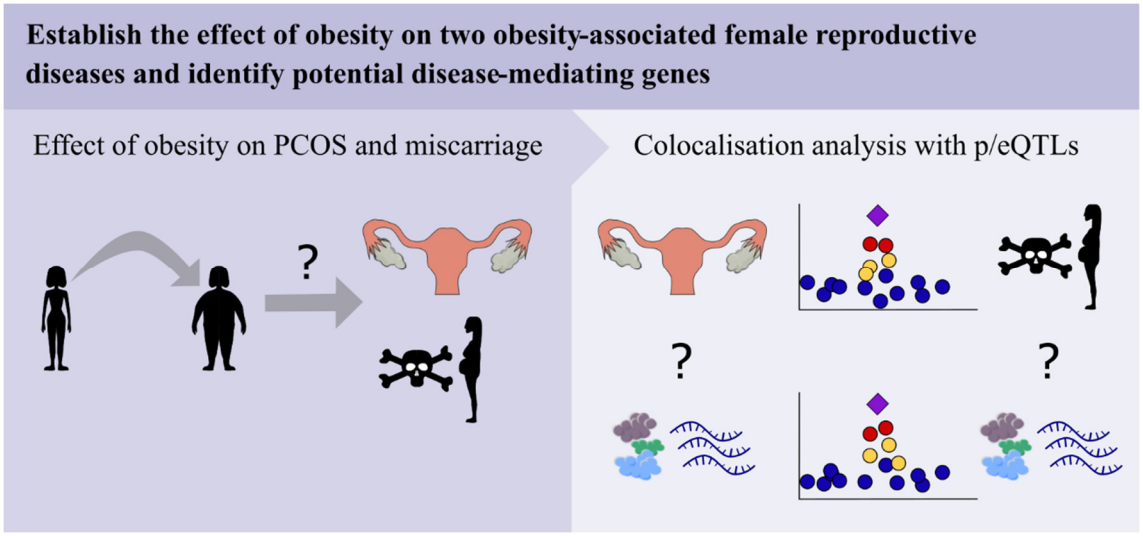


Figure 4.1. Graphical abstract of the primary research aim and research objectives for Chapter 4. In this chapter, my main research aim was to establish the effect of obesity on two female reproductive diseases and to identify disease-mediating genes for these diseases. To achieve this, my first research objective was to perform MR analysis of obesity traits on PCOS and miscarriage. My second research objective was to perform colocalisation analysis between these female reproductive traits and protein and expression levels.

4.2. Methods

4.2.1. Study overview

I first performed a bidirectional MR study of three obesity traits on PCOS and miscarriage (Davey Smith and Ebrahim, 2003). I had then planned to use eCAVIAR (Hormozdiari *et al.*, 2016) to conduct colocalisation analyses between the female reproductive diseases and genes/proteins, and therefore evaluated the software. However, the results of these evaluations raised concerns about

the reliability of eCAVIAR (see upcoming sections 4.2.4 and 4.3.2), and I therefore decided to use the colocalisation software Coloc, by Giambartolomei *et al.* (2014), instead. Finally, I performed a range of sensitivity analyses to ensure the robustness of my findings, as well as devised an exploratory new approach – Interaction-Coloc – that may aid in nuancing colocalisation results (Giambartolomei *et al.*, 2014).

The evaluation of eCAVIAR, the MR studies, and the colocalisation analyses included in this chapter were conducted at different stages of my DPhil, and the colocalisation analyses are included in two separate articles and a corresponding pre-print (Censin *et al.*, 2020, 2021; Laisk *et al.*, 2020). Therefore, there are some minor methodological differences between the analyses. The largest difference is which quantitative trait loci datasets I included in the colocalisation analyses (GTEx, INTERVAL and eQTLgen for PCOS, versus GTEx, INTERVAL and REPROMETA for recurrent miscarriage) (GTEx Consortium, 2017; Sun *et al.*, 2018; Võsa *et al.*, 2018; Kikas *et al.*, 2019). Any other methodological differences are minor and are clearly described in the methods section.

4.2.2. Datasets

4.2.2.1. Summary statistics for the female reproductive diseases

4.2.2.1.1. Polycystic ovary syndrome

Summary statistics for PCOS were taken from a case-control GWAS by Day *et al.* (2018). The case criteria varied between the different cohorts included in the study, but cases fulfilled at least one of the following set of criteria:

a) The National Institutes of Health criteria (Zawadzki and Dunaif, 1992; Day *et al.*, 2018), which requires that other aetiologies for the symptoms have been excluded, as well as both:
    i. Oligo- or anovulation.
    ii. Clinical or biochemical signs of hyperandrogenism.

b) The Rotterdam criteria (Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group, 2004; Day *et al.*, 2018), which requires that other aetiologies for the symptoms have been excluded, as well as at least two out of three of:
    i. Oligo- or anovulation (defined as oligo- or amenorrhea in some of the cohorts included in the study).
    ii. Clinical or biochemical signs of hyperandrogenism.
    iii. Polycystic ovaries.
c) Self-report of having PCOS (for the 23andMe cohort only) (Day *et al.*, 2018).

Up to 10,074 cases and 103,164 controls, all of European ancestry, had been included, and 14 independent loci surpassed the genome-wide association threshold ($P<5\times10^{-8}$; Table 4.1) (Day *et al.*, 2018). However, publicly available summary statistics were only available for:

a) **Top 10,000 SNPs dataset:** The top 10,000 SNPs with the lowest P-values, with association statistics based on all available cases and controls.
b) **Without-23 dataset:** All investigated SNPs, but with association statistics from analyses in which the largest cohort (23andMe) had been excluded and analyses thus only performed in up to 4,890 cases and 20,405 controls.

Table 4.1. The 14 genetic loci associated with PCOS.

SNP	Chromosome	Position	EA	NEA	EAF	OR (95% CI)	P
rs2178575	2	213391766	A	G	0.15	1.18 (1.13-1.23)	3.3×10^{-14}
rs11031005	11	30226356	C	T	0.15	1.17 (1.12-1.23)	8.7×10^{-13}
rs804279	8	11623889	A	T	0.26	1.14 (1.10-1.18)	3.8×10^{-12}
rs11225154	11	102043240	A	G	0.09	1.20 (1.13-1.26)	5.4×10^{-11}
rs9696009	9	126619233	A	G	0.07	1.22 (1.15-1.30)	8.0×10^{-11}
rs13164856	5	131813204	T	C	0.73	1.13 (1.09-1.18)	1.5×10^{-10}
rs1784692	11	113949232	T	C	0.82	1.15 (1.10-1.21)	1.9×10^{-10}
rs7563201	2	43561780	G	A	0.55	1.11 (1.08-1.15)	3.7×10^{-10}
rs8043701	16	52375777	T	A	0.18	1.14 (1.09-1.18)	9.6×10^{-10}
rs1795379	12	75941042	C	T	0.76	1.12 (1.08-1.17)	1.8×10^{-9}
rs853854	20	31420757	T	A	0.50	1.10 (1.07-1.14)	2.4×10^{-9}
rs2271194	12	56477694	A	T	0.42	1.10 (1.07-1.14)	4.6×10^{-9}
rs10739076	9	5440589	A	C	0.31	1.12 (1.07-1.16)	2.5×10^{-8}
rs7864171	9	97723266	G	A	0.57	1.10 (1.06-1.13)	3.0×10^{-8}

Summary statistics for the top SNPs in each associated loci as per Day *et al.* (2018). Genome position in GRCh 37. Table adapted (updated text) from Censin *et al.* (2021) and its corresponding pre-print (Censin *et al.*, 2020). NEA, non-effect allele.

Three different versions of these publicly available PCOS datasets were considered for use in the colocalisation analyses:

a) **Main dataset:** Combining the two publicly available datasets, with preference given to SNP statistics computed in the Top 10,000 SNPs dataset over the Without-23 dataset. Thus, this version of the PCOS dataset had the maximum sample size and number of included SNPs, but unequal sample sizes between SNPs.
b) **Top 10,000 SNPs dataset:** Using only the dataset in which the statistics had been computed using all available cases and controls (including the 23andMe cohort), thus having a limited set of SNPs, but the maximum and approximately similar sample size for all SNPs.
c) **Without-23 dataset:** Using only the dataset with the statistics available for all investigated SNPs from the analyses excluding the 23andMe cohort, thus having the maximum number of SNPs and an approximately similar (but limited) sample size for all SNPs.

4.2.2.1.2. Recurrent miscarriage

Dr. Laisk and collaborators conducted GWASs for sporadic and recurrent miscarriage in one of the articles underlying this thesis chapter (Laisk *et al.*, 2020). This GWAS had identified only a single risk locus (rs146350366) for sporadic miscarriage (Laisk *et al.*, 2020). Although my collaborators had found chromatin interactions with several genes for this SNP, the most interesting gene (*FGF9*) was too distant from rs146350366 for the required summary statistics to be available in the quantitative trait loci datasets (GTEx Consortium, 2017; Kikas *et al.*, 2019; Laisk *et al.*, 2020).

Recurrent miscarriage cases were defined as those with a diagnosis of "recurrent pregnancy loss" (ICD-10 code N.96) or that had self-reported three or more consecutive miscarriages (Laisk *et al.*, 2020). A GWAS was then performed in up to 750 cases and 150,215 controls of European ancestry which identified three independent genetic loci that passed quality control and surpassed the multiple testing threshold ($P<5\times10^{-8}$; Table 4.2) (Laisk *et al.*, 2020). Given the scope of the project, the lack of suitable quantitative trait loci data for sporadic miscarriage, and the more serious

phenotype of recurrent miscarriage, I decided to focus on recurrent miscarriage for the colocalisation analyses. The meta-analysis of the different cohorts had been performed using Stouffer's Z-score method, and thus only Z-scores were available for all SNPs (and not effect estimates and standard errors) (Willer, Li and Abecasis, 2010; Laisk *et al.*, 2020).

Table 4.2. The three genetic loci associated with recurrent miscarriage.

SNP	Chromosome	Position	EA	NEA	EAF	OR (95% CI)	P
rs143445068	11	19632396	C	G	0.01	3.4 (2.4-5.0)	5.2×10^{-9}
rs7859844	9	83677992	A	G	0.06	1.7 (1.4-2.0)	1.3×10^{-8}
rs183453668	21	44831990	A	G	0.01	3.8 (2.4-5.9)	2.8×10^{-8}

Summary statistics for the top SNPs in each loci as per Laisk *et al.* (2020). Genome position in GRCh 37. Summary statistics only provided in Z-scores and information on CIs thus only provided to the first decimal place, as in the original study (Laisk *et al.*, 2020).

4.2.2.1.3. Miscarriage and stillbirth

Genome-wide summary statistics for "miscarriage and stillbirth" from the UKBB analyses by Zhou *et al.* (2018) were used for the MR analyses. These analyses had been performed in up to 5,463 cases and 400,946 controls (male and female) of White British ancestry using Scalable and Accurate Implementation of GEneralized mixed model (SAIGE) (Zhou *et al.*, 2018). Cases had been identified using PheCode 634 ("miscarriage and stillbirth"), which had been derived from a combination of ICD-9 codes, including ICD-9 code 634 ("Spontaneous abortion"), 634.0 ("Spontaneous abortion complicated by genital tract and pelvic infection"), 631.0 ("Inappropriate change in quantitative human chorionic gonadotropin (hCG) in early pregnancy"), as well as 631 and 631.8 ("Other abnormal product of conception") (Denny *et al.*, 2013; Zhou *et al.*, 2018). In addition, Zhou *et al.* (2018) also identified cases using ICD-10 codes that were mapped to each PheCode using, for example, maps from Unified Medical Language System (National Library of Medicine, 2021), string matching, other available resources and manual review.

4.2.2.2. Summary statistics for expression and protein quantitative trait loci datasets

4.2.2.2.1. INTERVAL

The INTERVAL study consists of approximately 50,000 blood donors recruited between 2012 and 2014, for details about the study see section 3.2.2.1 (Di Angelantonio *et al.*, 2017). For a subset of the study, protein levels had been quantified using the SomaLogic assay (Di Angelantonio *et al.*, 2017; Sun *et al.*, 2018). Genome-wide summary statistics have been made publicly available for 3,283 aptamers measuring 2,994 unique proteins in 3,301 participants of European ancestry (Sun *et al.*, 2018).

4.2.2.2.2. GTEx

Genotype-expression data from version 7 of the GTEx project were used (GTEx Consortium, 2017). Version 7 of GTEx contains publicly available genotype-expression summary statistics for 48 tissues, with sample sizes ranging between 80 to 491 for each tissue and approximately 85% of donors being of European ancestry (GTEx Consortium, 2017, 2020). Summary statistics for SNP associations with gene expression levels are available for genetic variants within 1 Mb of the transcription start site of each gene (GTEx Consortium, 2017, 2020). GTEx allele frequency information was extracted from http://cnsgenomics.com/software/smr/#DataResource as needed (Zhu *et al.*, 2016). For the miscarriage colocalisation analyses, one SNP (rs144887114) in the terminal ileum dataset was excluded in the colocalisation analyses of *RP11-154D17.1* and *TLE1* due to having improbable effect estimates of over a trillion SD per EA (GTEx Consortium, 2017, 2020).

4.2.2.2.3. eQTLgen

The eQTLgen consortium is a large blood eQTL analysis in up to 31,684 individuals of primarily of European ancestry (Võsa *et al.*, 2018). Data of both cis genotype associations (within 1 Mb of the centre of each gene) and trans genotype associations (over 5 Mb from the centre of each gene) are publicly available (Võsa *et al.*, 2018).

4.2.2.2.4. REPROMETA

REPROMETA is a cis-eQTL study in placenta from 40 women of European decent, with association statistics publicly available for SNPs within 100 kb of each gene (Kikas *et al.*, 2019). As it was unclear what genome build the rsIDs had been mapped to, genome positions were updated according to GRCh 37 using biomaRt to ensure correct mapping between datasets (Durinck *et al.*, 2005, 2009).

4.2.2.3. Summary statistics for the obesity traits

GWAS summary statistics for BMI, WHR, and WHRadjBMI were taken from the GIANT consortium (Locke *et al.*, 2015; Shungin *et al.*, 2015; Pulit *et al.*, 2019). Statistics based on analyses both including and excluding the UKBB were used in the MR analyses, and all data were from analyses of women of European descent (Locke *et al.*, 2015; Shungin *et al.*, 2015; Pulit *et al.*, 2019). The total sample size in each GWAS was 171,977 and 434,794 women for BMI, 118,004 and 381,152 women for WHR, and 116,742 and 379,501 women for WHRadjBMI, for the analyses in GIANT only and GIANT and UKBB meta-analysis, respectively (Locke *et al.*, 2015; Shungin *et al.*, 2015; Pulit *et al.*, 2019).

4.2.2.4. UK Biobank

Briefly, the UKBB had been centrally processed as outlined in section 2.2.2 (Sudlow *et al.*, 2015; Bycroft *et al.*, 2018). I then performed sample and post-imputation quality control using similar criteria as in section 2.2.2, but restricting my sample to the "White British ancestry" subset and not including any samples related to a third degree or higher (kinship coefficient>0.0442) (Manichaikul *et al.*, 2010; Bycroft *et al.*, 2018). For LD-reference panels, I then randomly extracted 3,927 female samples for the MR analyses, 29,454 women for the PCOS colocalisation analyses, and 27,625 participants for the recurrent miscarriage colocalisation analyses. The imputed data were then converted to hard-calls (--hard-call-threshold 0.1) (Purcell *et al.*, 2007). Analyses were performed in PLINK versions 1.90b3 and 2.00a-20170724 (Purcell *et al.*, 2007).

The UKBB has a Research Tissue Bank approval, all participants gave informed consent, and this study was performed under UKBB study application ID 11867 (Sudlow *et al.*, 2015).

4.2.2.5. 1000 Genomes

An LD-reference panel using data from 1000 Genomes phase 3 was used for the initial eCAVIAR investigations (Auton *et al.*, 2015). Only participants of European ancestry were included, and one individual in each pair related to a third degree or closer were excluded, resulting in a final sample size of 503 samples (Auton *et al.*, 2015). Genotype quality control was then performed by excluding SNPs with HWE $P<1\times10^{-6}$, missing call rate >0.05, and minor allele frequency <1% using PLINK v1.90b3 (Purcell *et al.*, 2007), similar to previous studies in the 1000 Genomes dataset (Broad, 2014; Auton *et al.*, 2015; Pers, Timshel and Hirschhorn, 2015; Berg *et al.*, 2019).

4.2.2.6. Common quality control procedures for all datasets

For all summary statistics datasets, SNPs were excluded if they were non-biallelic, duplicated, indels, or were missing relevant data. SNP information, rs-aliases, rsIDs and genome position in GRCh 37 coordinates were taken from the biomaRt R package (Durinck et al., 2005, 2009) and from the NCBI as needed (using the files "All_20180423.vcf.gz" from ftp://ftp.ncbi.nih.gov/snp/organisms/human_9606_b151_GRCh37p13/VCF/ and "RsMergeArch.180920.txt.gz" from ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606/database/organism_data/) (Sherry et al., 2001).

4.2.2.7. Deciding which genes and proteins to investigate in each dataset for the colocalisation analyses

The e/pQTL datasets provided association statistics for different region sizes around each gene/protein (depending on for example distance from transcription start site or centre of gene) (GTEx Consortium, 2017; Sun *et al.*, 2018; Võsa *et al.*, 2018; Kikas *et al.*, 2019). For Coloc, a region size of 400 kb centred on the top disease SNP is recommended, whereas only the 100 SNPs closest to the top disease SNP is recommended for eCAVIAR (Giambartolomei *et al.*, 2014;

Hormozdiari *et al.*, 2016). To ensure that association statistics were available for the full 400 kb region recommended for Coloc, genes were only included if their transcription start site or centre was within 800 kb of each disease association peak for GTEx and eQTLgen, respectively (GTEx Consortium, 2017; Võsa *et al.*, 2018). Genome-wide association statistics were available for INTERVAL (Sun *et al.*, 2018). However, I only investigated proteins that consisted of a gene with a transcription start site within 800 kb of the disease association peaks to have a similar methodology to the eQTL datasets in the PCOS analyses (Durinck *et al.*, 2005, 2009; Sun *et al.*, 2018). For the miscarriage colocalisation analyses, I investigated all proteins consisting of a gene within 5 Mb of each miscarriage index variant. For REPROMETA, only genes with association statistics for SNPs within 200 kb of the top miscarriage SNPs were considered, since summary statistics were only available for genetic variants within 100 kb of each gene in that dataset (Kikas *et al.*, 2019).

In the original eCAVIAR article (Hormozdiari *et al.*, 2016), only genes which harboured a quantitative trait locus were considered. Therefore, proteins in INTERVAL were only brought forward for analysis with eCAVIAR if at least one SNP in the investigated region was associated with the protein at an FDR-corrected P-value <0.05 (computed in the summary GWAS file for each protein) (Sun *et al.*, 2018). For colocalisation analyses involving the eQTLgen dataset, only genes which had at least one SNP in the region that was associated with the gene at a permuted FDR <0.05 (computed by eQTLgen) were analysed (Võsa *et al.*, 2018).

4.2.3. Mendelian randomisation analyses

Bidirectional MR analyses of BMI, WHR, and WHRadjBMI on PCOS and miscarriage were performed (Davey Smith and Ebrahim, 2003). The top independent SNPs (r^2 threshold of 0.01) associated with each trait at $P<5\times10^{-8}$ were identified using --clump in PLINK v1.90b3 (Purcell *et al.*, 2007). Thereafter, the LD between all pairs of SNPs included in an instrument was computed using PLINK v1.90b3 (Purcell *et al.*, 2007). In the case of long-distance LD (r^2 threshold of 0.01),

the SNP with the highest P-value in each correlated pair was removed. Palindromic SNPs were excluded unless the minor allele frequency was <0.4 in both the exposure and the outcome dataset, and the minor allele was the same in both datasets (Burgess and Thompson, 2015; Burgess, Davey Smith, *et al.*, 2020). If SNPs were missing in any dataset, correlated proxies at $r^2 \geq 0.8$ and within 250 kb were used instead, if available (Burgess *et al.*, 2015; Adams and Neuhausen, 2019; Burgess, Davey Smith, *et al.*, 2020).

For analyses involving PCOS, obesity summary statistics from the analyses where the UKBB had been included were used (Day *et al.*, 2018; Pulit *et al.*, 2019). MR analyses (including IVW, median-based methods, and MR-Egger) were then performed using the R package "MendelianRandomization" (Bowden, Davey Smith and Burgess, 2015; Burgess *et al.*, 2015; Bowden *et al.*, 2016; Yavorska and Burgess, 2017).

The GWAS on recurrent miscarriage by Laisk *et al.* (2020) only provided Z-scores, instead of effect estimates and standard errors. To investigate if there was any evidence of obesity affecting the risk of miscarriage, the summary statistics from the GWAS on miscarriage and stillbirth by Zhou *et al.* (2018) were used instead. Otherwise, analyses were performed using the same procedure as outlined above and using data from the obesity GWASs excluding the UKBB to avoid the risk of overfitting and weak instrument bias affecting the results, albeit such a risk should be negligible (see section 1.4.2.8.1) (Burgess and Thompson, 2011; Taylor *et al.*, 2014; Locke *et al.*, 2015; Shungin *et al.*, 2015; Burgess, Davies and Thompson, 2016). To investigate if recurrent miscarriage had an effect on obesity, a fixed-effect meta-analysis of the outcome summary statistics for the miscarriage SNPs was performed using the R-package metafor (Davey Smith and Ebrahim, 2003; Viechtbauer, 2010; Burgess and Thompson, 2015). However, this approach will only test if there is evidence of a causal effect, and not provide any causal effect estimates (Davey Smith and Ebrahim, 2003; Burgess and Thompson, 2015).

4.2.4. Primary evaluation of eCAVIAR

4.2.4.1. Initial eCAVIAR evaluations

I initially evaluated eCAVIAR for use in the colocalisation analyses (Hormozdiari *et al.*, 2016). Since the recurrent miscarriage dataset provided Z-scores rather than effect estimates and standard errors, I focused on PCOS in these initial evaluations for simplicity. Due to the large scale of the GTEx dataset, my initial evaluations of eCAVIAR also focused on pQTL data from INTERVAL and eQTL data from eQTLgen (GTEx Consortium, 2017; Sun *et al.*, 2018; Võsa *et al.*, 2018).

4.2.4.2. eCAVIAR settings

4.2.4.2.1. Polycystic ovary syndrome dataset

Analyses were performed using all three PCOS datasets, including the top 10,000 SNPs dataset (including the 23andMe cohort, but with a limited number of SNPs), the Without-23 dataset (excluding the 23andMe cohort, but with the maximum number of SNPs), and the main PCOS dataset (a combined version of the two former datasets, with preference given to summary statistics computed from the analyses including the 23andMe cohort) (Day *et al.*, 2018).

4.2.4.2.2. Region size

Different region sizes have been used in previous studies using eCAVIAR (Hormozdiari *et al.*, 2016; Aylward *et al.*, 2018; B. Liu, Calton, *et al.*, 2018; B. Liu, Pjanic, *et al.*, 2018). Initially, I had planned to use the recommended region size of +/- 50 SNPs of the top disease SNP (Hormozdiari *et al.*, 2016). However, closer inspection of disease risk loci indicated that such a narrow region size risked not including the entire peak of associated SNPs (see, for example, Appendix Fig 4.1). After personal correspondence with eCAVIAR originator Dr. Hormozdiari (2019, e-mail, February 27) about how best to proceed, the initial eCAVIAR evaluations were performed using a larger region size of +/- 100 kb of the top PCOS hit.

4.2.4.2.3. Maximum number of causal variants

Computation time increased exponentially with number of SNPs and higher maximum number of causal variants. Therefore, analyses were performed supplying three different values for "maximum number of causal SNPs" (1, 2, and 3), as higher maximum number of causal variants were computationally infeasible to evaluate for many genes/proteins.

4.2.4.2.4. Colocalisation threshold

Genes and proteins were considered to be colocalising with risk of PCOS if any SNP had a CLPP score >0.01, as in the original eCAVIAR article (Hormozdiari *et al.*, 2016).

4.2.5. Main colocalisation analyses performed using Coloc

4.2.5.1. Explorative analysis of the impact of only including the SNPs most robustly associated with polycystic ovary syndrome

I was uncertain if having very unequal sample sizes for different SNPs would affect colocalisation results. In an attempt to reduce the potential of such bias and to maximise power, I first performed the Coloc analyses using the same settings as detailed below but using the dataset that was based on the 10,000 SNPs most robustly associated with PCOS.

4.2.5.2. Coloc settings

4.2.5.2.1. Polycystic ovary syndrome dataset

The main PCOS dataset was used in the main Coloc analyses, using all available SNPs and with effect estimates preferentially taken from the analyses including the 23andMe cohort.

4.2.5.2.2. Colocalisation method in Coloc

Approximate Bayes factor colocalisation was used to investigate colocalisation, using the coloc.abf() function (Giambartolomei *et al.*, 2014).

4.2.5.2.3. Region size

Previous studies have used different region sizes when investigating colocalisation using Coloc, commonly 400 kb and 2 Mb region sizes (Guo *et al.*, 2015; Franceschini *et al.*, 2018; Çaliskan *et al.*, 2019; Raulerson *et al.*, 2019). In the main analyses, a region size of 2 Mb was used. Due to data limitations, this region size was only centred around the top PCOS/miscarriage SNP in INTERVAL (Sun *et al.*, 2018). In GTEx and eQTLgen, this region size included all available SNPs for each gene and with SNPs spanning at least 200 kb on each side of the top PCOS/miscarriage SNP (GTEx Consortium, 2017; Võsa *et al.*, 2018). For REPROMETA (Kikas *et al.*, 2019), there was only one gene (*NAV2*) with genotype information covering 200 kb around a miscarriage index variant (rs143445068). However, genotype associations for *NAV2* spanning the full 2 Mb were unavailable, wherefore all the SNPs available in REPROMETA were used in the analyses spanning 2 Mb (Kikas *et al.*, 2019).

4.2.5.2.4. Choice of priors

The standard priors for p1 and p2 ($p1=p2=1\times10^{-4}$) were used, as a previous simulation study have indicated these settings to be reasonable under most scenarios (Giambartolomei *et al.*, 2014; Wallace, 2020). However, a stricter prior ($p12=1\times10^{-6}$) than default ($p12=1\times10^{-5}$) was used for p12, as the default prior has been shown to be too lenient in many scenarios (Giambartolomei *et al.*, 2014; Wallace, 2020).

4.2.5.2.5. Colocalisation threshold

Genes and proteins with a PP≥0.75 of colocalising with PCOS or miscarriage were regarded as having strong evidence of colocalisation (Franceschini *et al.*, 2018). However, given the hypothesis-generating approach of the research, genes and proteins with a PP>0.50 of colocalisation were regarded as having nominal evidence of having a shared causal variant with the reproductive diseases (Donovan *et al.*, 2020; Gay *et al.*, 2020; Oliva *et al.*, 2020). Of note, slightly different thresholds were used to denote evidence of colocalisation in the article underlying the miscarriage analyses (suggestive evidence: PP>0.5; evidence: PP>0.8, strong evidence: PP>0.9)

(Laisk *et al.*, 2020), but this discrepancy between the article and the chapter had no impact on the interpretation of the miscarriage colocalisation results.

4.2.5.2.6. Power threshold

Power for detecting colocalisation was computed as the sum of PPs for the hypotheses of no colocalisation (H3) and colocalisation (H4) (Giambartolomei *et al.*, 2014; Çaliskan *et al.*, 2019). The threshold for determining if an analysis had sufficient power to determine colocalisation was set to 0.80 (Çaliskan *et al.*, 2019).

4.2.5.2.7. Conditional analyses for recurrent miscarriage

All colocalisation analyses where the top available miscarriage SNP was associated with the gene expression or protein levels at $P<0.05$ were visually inspected using LocusCompareR (Liu, Gloudemans and Montgomery, 2018) to investigate if there were any regions with possible allelic heterogeneity. In one region, visual inspection of the rs7859844 miscarriage locus and *RP11-154D17.1* expression in visceral adipose tissue indicated that the region might contain two independent eQTLs. As Coloc assumes a single causal variant in each locus (Giambartolomei *et al.*, 2014), the colocalisation analysis was reconducted using effect estimates conditioned on the top hit. The top eQTLs were identified using GCTA version 1.91.4 (--cojo-top-SNPs) (Yang *et al.*, 2011, 2012). Effect estimates conditioned on the top SNP using GCTA (--cojo-cond) were then computed before reconducting the colocalisation analyses using the same approach as in the main analysis (Yang *et al.*, 2011, 2012).

4.2.5.2.8. Additional Coloc settings

Coloc requires user-supplied input of both total sample size and the number of cases. For analyses involving the main or top 10,000 SNPs PCOS datasets, the number of cases was set to 10,074 and the total sample size set to 113,238 (Day *et al.*, 2018). For the sensitivity analyses using the PCOS dataset where the 23andMe cohort had been excluded, the number of cases was set to 4,890 and the total sample size set to 25,295 (Day *et al.*, 2018). For GTEx, INTERVAL, and REPROMETA, the sample size was set to that reported for each tissue, and for eQTLgen the average sample size for

the included genetic variants was used (GTEx Consortium, 2017; Sun *et al.*, 2018; Võsa *et al.*, 2018; Kikas *et al.*, 2019). Minor allele frequencies from the relevant reproductive disease dataset were used in all Coloc analyses (Day *et al.*, 2018).

4.2.6. Sensitivity analyses

4.2.6.1. Reperforming the colocalisation analyses using the polycystic ovary syndrome dataset where the 23andMe cohort had been excluded

The Coloc analyses were also conducted using the PCOS dataset where the 23andMe cohort had been excluded to ensure similar sample sizes for all genetic variants, in case unequal sample sizes for different SNPs might bias colocalisation results (Day *et al.*, 2018).

4.2.6.2. Region sizes

In addition to evaluating the evidence of colocalisation using 2 Mb regions around the index SNPs, PPs were also computed for two additional region sizes as a sensitivity analysis:

a) **400 kb region:** Centred around the top PCOS/miscarriage SNP and performed in all e/pQTL datasets.

b) **Independent region:** Including all genetic variants in the top PCOS SNPs "independent region" (approximately independent regions of LD in Europeans) (Berisa and Pickrell, 2016). Analyses using this region size were only performed in INTERVAL due to data limitations in the other datasets, and only for colocalisation analyses involving PCOS.

4.2.6.3. HyPrColoc

As a sensitivity analysis, colocalisation was also evaluated using HyPrColoc (Foley *et al.*, 2021), a recently developed extension of Coloc (Giambartolomei *et al.*, 2014). The added value of HyPrColoc is that it allows for evaluation of more than two traits in a single analysis (Foley *et al.*, 2021). However, when this work was conducted the HyPrColoc article has not undergone peer-review yet and was therefore only considered as a sensitivity analysis (Foley *et al.*, 2021).

The default priors of prior.1=1×10^{-4} and prior.2=0.98 were used in the HyPrColoc analyses (Foley *et al.*, 2021). However, the alignment and regional probability thresholds were set to 0.8, which is more stringent than the default settings (Foley *et al.*, 2021).

Since only Z-scores were available in the eQTLgen dataset, betas and standard errors were estimated with the following formulas:

$$\hat{\beta} = \frac{z}{\sqrt{2p(1-p)(n+z^2)}}$$

$$Standard\ error = \frac{1}{\sqrt{2p(1-p)(n+z^2)}}$$

where z is the Z-score, p is the minor allele frequency and n is the sample size (Zhu *et al.*, 2016; Võsa *et al.*, 2018).

4.2.6.4. Interaction-Coloc - protein-protein interaction follow-up analyses using Coloc

4.2.6.4.1. Rationale

As evident from the literature, genes can share regulatory features, which can cause several genes to colocalise with a trait at a single locus (Bhatia and Kleinjan, 2014; Hormozdiari *et al.*, 2016; Krijger and De Laat, 2016; Yao *et al.*, 2017). Since the aim of this chapter was to highlight potential disease-mediating genes/proteins, I wanted to see if it was possible to nuance the results for the individual genes colocalising at the same loci. I thus investigated a new approach to use Coloc (Giambartolomei *et al.*, 2014), which I call "Interaction-Coloc".

In brief, I looked to see if genes/proteins that are linked (via protein-protein interactions) to the genes/proteins identified in the main analysis gave further evidence for a role of their involvement in disease pathophysiology. In a hypothetical example, illustrated in Fig 4.2, two genes may colocalise with disease risk at a certain locus due to common regulatory features, even if only one of them has a true effect on the disease. We can then investigate if other genes – affiliated with the first genes but in another part of the genome – have evidence of colocalisation with the disease. If

one of the affiliated genes also have evidence of colocalisation, I reasoned that this should strengthen the evidence of that pair and their affiliated pathway mediating the disease risk.

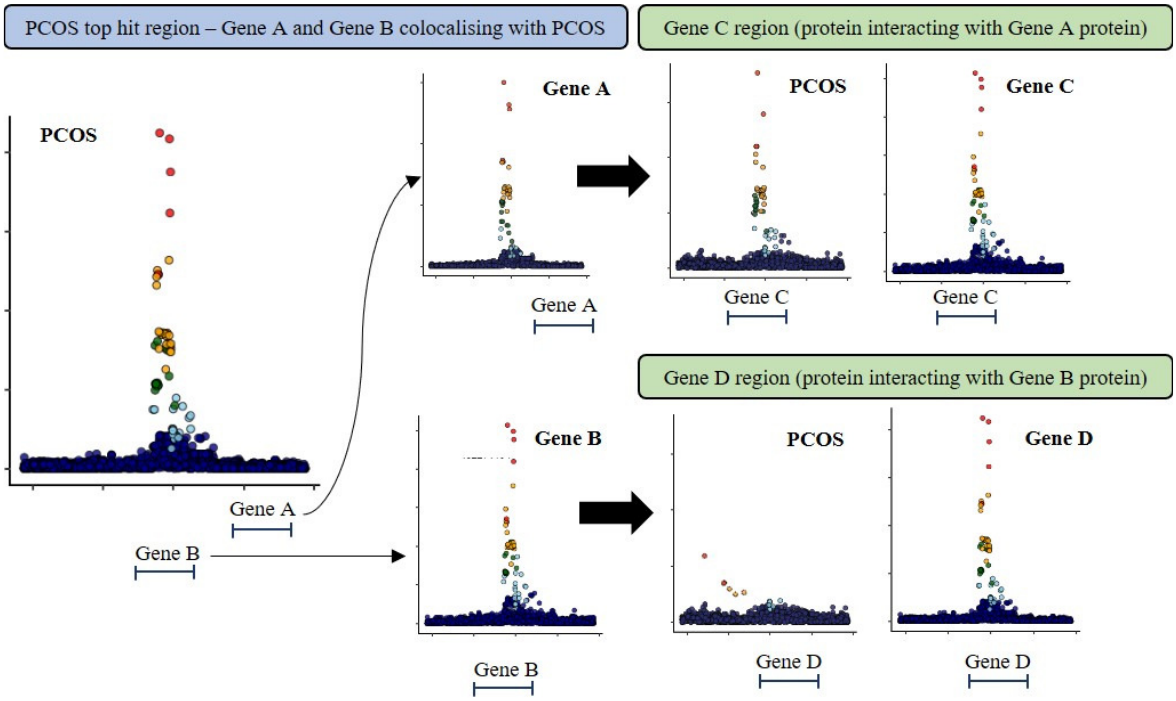


Figure 4.2. Illustration of the Interaction-Coloc rationale. In a hypothetical scenario, Gene A and B are both colocalising with risk of PCOS in a certain part of the genome due to shared regulatory features. To nuance the evidence of which gene may be the true mediating gene, we can use protein-protein interaction data to find other genes/proteins that are linked to Genes A and B – in this case Gene C (protein interacting with Gene A protein) and Gene D (protein interacting with Gene B protein). If there is evidence of colocalisation for Gene C but not Gene D with PCOS risk at those genes' respective genomic location, this should in theory strengthen the evidence of Gene A (and its affiliated pathways) being the mediating gene at the original disease-associated locus. Note that the PCOS-association at the genomic location of Gene C and D would not be expected to surpass the genome-wide significance threshold, as the loci then would have been investigated in the primary colocalisation analyses. Figure adapted (changed text) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

4.2.6.4.2. Method

Protein-protein interaction data in humans from Reactome were used to identify proteins that were "linked" to the originally identified genes and proteins (Fabregat *et al.*, 2018). I extracted all genes that encode proteins linked to any of the genes/proteins identified in the main colocalisation analysis. Only genes with Ensembl gene identifiers were kept. For FSH, only interactions listed for the beta subunit were included, since the alpha subunit also forms part of other hormones (National Center for Biotechnology Information, 2020). Mapping between different datasets was done using GRCh 37 and biomaRt as needed, and only transcripts with a numeric autosomal chromosome were included to ensure correct mapping between datasets (Durinck *et al.*, 2005, 2009).

Colocalisation was then performed using Coloc (Giambartolomei *et al.*, 2014). All SNPs within 1 Mb of the average transcription start site were included in the analyses for INTERVAL (Sun *et al.*, 2018). For the other datasets all available SNPs were used, resulting in 2 Mb region sizes for all datasets.

Since the genes and proteins in the Interaction-Coloc analyses by definition were previously linked to genes/proteins suspected to be involved in the diseases, the default Coloc p12 prior ($p12=1\times10^{-5}$) was used for these analyses (Giambartolomei *et al.*, 2014). However, for completeness and as a sensitivity analysis, the Interaction-Coloc analyses were also performed using the same p12 setting as in the main analysis ($p12=1\times10^{-6}$).

4.2.6.4.3. Colocalisation analysis of conditioned variants for FSHR

Visual inspection of one of the Interaction-Coloc analyses suggested that there were two independent eQTLs for *FSHR* expression in testis. To determine which eQTL that was driving the colocalisation, colocalisation analysis was performed using effect estimates conditioned on each of the top independent variants in the region.

The top independent variants in the *FSHR*-region were identified using the --clump command in PLINK (Purcell *et al.*, 2007). As the genetic variants were not genome-wide significant, they were identified using a --clump-p1 setting of 0,001, --clump-r2 setting of 0.05, and performing the clumping on the entire locus (using the UKBB LD-reference panel constructed for the PCOS analyses) (Purcell *et al.*, 2007). Two independent SNPs (rs4953650 and rs2349415) were identified. Conditional analyses for each of these two SNPs was then performed using GCTA version 1.91.4 in both the *FSHR* eQTL dataset (in testis), as well as the main PCOS dataset (Yang *et al.*, 2011, 2012).

4.2.7. Phenome-wide association study and in-silico investigations

In addition to the colocalisation analyses, an *in-silico* follow-up was performed of all the index SNPs associated with a female reproductive disease that also had evidence of colocalisation. First, the top hit's associations with other traits and diseases were assessed in phenome-wide association data from the Open Targets Genetics platform (Carvalho-Silva *et al.*, 2019). The significance threshold for these analyses was set to $P<1.7\times10^{-6}$ (as visual inspection of the Bonferroni-corrected threshold on the Open Targets platform (Carvalho-Silva *et al.*, 2019) suggested a threshold of approximately $P<1\times10^{-5}$, then further correction for six SNPs resulting in $P<1.7\times10^{-6}$). Secondly, the evidence of a regulatory role for these SNPs in various cell- and tissue types was assessed using the HaploReg resource (Ward and Kellis, 2012).

4.2.8. Data availability and software

4.2.8.1. Data availability

For data availability see Appendix Table 1.1.

4.2.8.1. Software

Analyses and plots were done using R versions 3.5.1 and 3.4.3 (R Core Team, 2014), bash version 4.1.2(2) and awk (Free Software Foundation, 2020), and R packages coloc (Giambartolomei *et al.*, 2014), hyprcoloc (Foley *et al.*, 2021), LocusCompareR (Liu, Gloudemans and Montgomery, 2018), tidyr (Wickham and Henry, 2018), data.table (Dowle and Srinivasan, 2020), biomaRt (Durinck *et al.*, 2005, 2009), plyr (Wickham, 2011), devtools (Wickham, Hester and Chang, 2020), and ggplot2 (Wickham, 2016).

4.3. Results

4.3.1. Mendelian randomisation analysis confirms a role of obesity in polycystic ovary syndrome, but not in miscarriage

There was strong evidence of BMI increasing the risk of PCOS in the main IVW MR analysis (OR 2.58 per 1 SD higher BMI, 95% CI 2.17-3.06, $P=1.5\times10^{-27}$; multiple testing threshold set to $P<0.008$ for the MR analyses, for 0.05/6 trait-disease combinations) (Bonferroni, 1936; Burgess and Thompson, 2015; Burgess *et al.*, 2015). A causal effect was also strongly supported by the sensitivity analyses using the median-weighted and MR-Egger methods (simple median $P=1.2\times10^{-7}$, weighted median $P=3.2\times10^{-6}$, MR-Egger $P=7.4\times10^{-10}$; Table 4.3) (Bowden, Davey Smith and Burgess, 2015; Bowden *et al.*, 2016). In addition, there was some evidence of an effect of WHR on PCOS risk in the main analysis (OR 1.45 per 1 SD higher WHR, 95% CI 1.19-1.75, $P=1.7\times10^{-4}$). The point estimates were directionally consistent in all the sensitivity analyses for both WHR and WHRadjBMI, but without the results surpassing the multiple testing threshold. P-values from the MR-Egger intercept test gave no strong evidence of directional pleiotropy (MR-Egger intercept P-values 0.05, 0.51, and 0.99 for BMI, WHR, and WHRadjBMI, respectively).

Table 4.3. MR results for the effect of obesity on PCOS.

Exposure	N SNPs analysis (proxies)	Method	OR (95% CI)	P
BMI	489 (3)	IVW	2.58 (2.17-3.06)	1.5×10^{-27}
BMI	489 (3)	Simple median	2.02 (1.56-2.63)	1.2×10^{-7}
BMI	489 (3)	Weighted median	2.49 (1.70-3.66)	3.2×10^{-6}
BMI	489 (3)	MR-Egger	3.85 (2.51-5.91)	7.4×10^{-10}
WHR	361 (6)	IVW	1.45 (1.19-1.75)	1.7×10^{-4}
WHR	361 (6)	Simple median	1.23 (0.94-1.61)	0.14
WHR	361 (6)	Weighted median	1.17 (0.88-1.54)	0.29
WHR	361 (6)	MR-Egger	1.25 (0.79-2.00)	0.34
WHRadjBMI	454 (6)	IVW	1.11 (0.96-1.29)	0.16
WHRadjBMI	454 (6)	Simple median	1.10 (0.88-1.38)	0.40
WHRadjBMI	454 (6)	Weighted median	1.10 (0.87-1.40)	0.42
WHRadjBMI	454 (6)	MR-Egger	1.11 (0.81-1.53)	0.52

Effect estimates per 1 female-specific SD higher obesity trait.

There was no evidence of an effect of the obesity traits on miscarriage and stillbirth, nor any evidence of directional pleiotropy in the MR-Egger intercept test (P>0.05 in all analyses) (Appendix Table 4.1). There was also no evidence of the reproductive diseases having a causal effect of obesity, nor any evidence of directional pleiotropy in these analyses (MR-Egger intercept test P>0.05) (Appendix Tables 4.2-4.3).

4.3.2. eCAVIAR results are sensitive to choice of settings

4.3.2.1. Overview of the eCAVIAR colocalisation results

The eCAVIAR results seemed to be heavily dependent on choice of PCOS dataset and/or number of SNPs included in the analysis, as well as the setting of maximum number of causal SNPs (Fig 4.3-4.4 and Appendix Fig 4.2). In general, it seemed that most analyses surpassing the standard colocalisation threshold of at least one SNP with a CLPP score >0.01 included fewer SNPs and/or had a higher maximum number of causal SNPs setting (Hormozdiari *et al.*, 2016). This pattern seemed particularly strong for the analyses using only the 10,000 SNPs most robustly associated with PCOS (Fig 4.3), but with similar results for the other two datasets (Fig 4.4. and Appendix Fig 4.2).

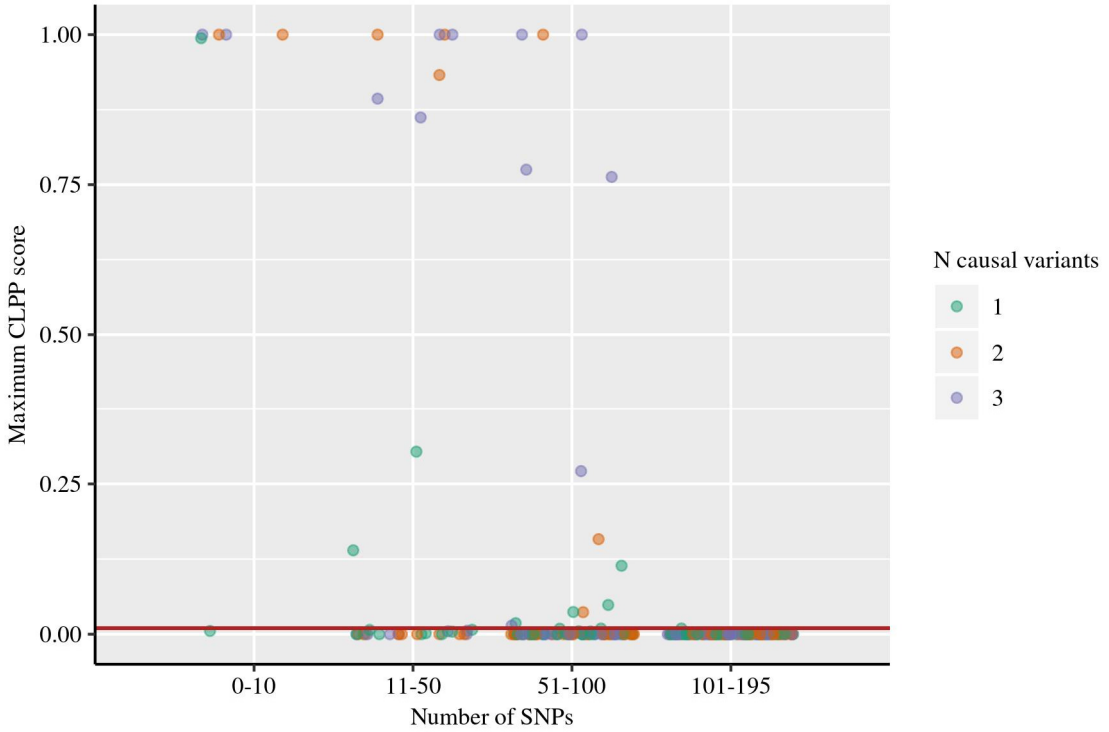


Figure 4.3. eCAVIAR colocalisation results for the top 10,000 SNPs PCOS dataset and gene expression and protein level data from GTEx and INTERVAL, respectively. Number of SNPs is the number of SNPs in the investigated region (+/- 100 kb from the top PCOS SNP) and maximum CLPP score the highest CLPP score of any analysed SNP in the investigated region. The red line marks the colocalisation threshold of a SNP with a CLPP score >0.01.

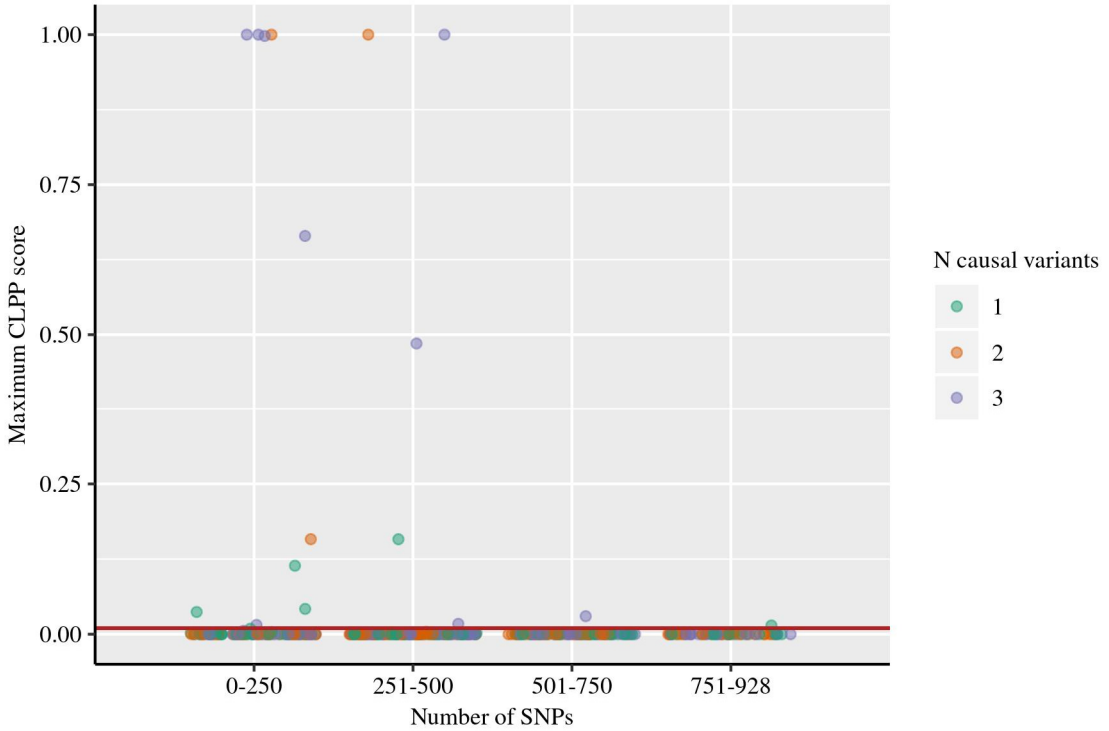


Figure 4.4. eCAVIAR colocalisation results using the main PCOS dataset and gene expression and protein level data from GTEx and INTERVAL, respectively. Number of SNPs is the number of SNPs in the investigated region (+/- 100 kb from the top PCOS SNP) and

maximum CLPP score the highest CLPP score of any analysed SNP in the investigated region. The red line marks the colocalisation threshold of a SNP with a CLPP score >0.01.

4.3.2.2. eCAVIAR colocalisation results seem very dependent on settings

To further understand the pattern of how the CLPP scores varied depending on the settings, I closely examined the PCOS loci and genes/proteins that eCAVIAR indicated were colocalising (CLPP>0.01) in any analysis (Table 4.4) (Hormozdiari *et al.*, 2016). These results indicated that the software computed very different maximum CLPP scores depending on PCOS dataset and maximum number of causal variants. Visual inspection of some of these loci indicated that eCAVIAR potentially yielded both false positive and false negative results. For example, the results suggested colocalisation of *RPS26* expression levels using the main PCOS dataset, but only when setting the maximum number of causal variants to one *or* three – with no evidence when setting the maximum number of causal variants to two (Table 4.4). However, visual inspection of the locus suggested that disease risk and *RPS26* blood expression levels were indeed colocalising (Appendix Fig 4.1). On the contrary, eCAVIAR provided evidence in favour of colocalisation between PCOS risk and *ORMDL2* expression levels using the main PCOS dataset and setting the maximum number of causal variants to three (Table 4.4). Yet, visual inspection of the locus did not suggest any colocalisation, and the top disease-associated variant rs2271194 was not even nominally associated with *ORMDL2* expression levels (Appendix Fig 4.3). Given the sensitivity of the eCAVIAR results to choice of settings and several examples of possible false positive and false negative results, I decided not to use the software in any further analyses and to disregard these results.

Table 4.4: eCAVIAR colocalisation results for all PCOS SNP and gene/protein combinations with evidence of colocalisation using any PCOS dataset and/or maximum number of causal variant setting.

Analysis			Top 10,000 SNPs PCOS Dataset			Excluding 23andMe PCOS Dataset			Main PCOS Dataset		
PCOS SNP	**Gene/Protein**	**QTL dataset**	**N=1**	**N=2**	**N=3**	**N=1**	**N=2**	**N=3**	**N=1**	**N=2**	**N=3**
rs10739076	*PDCD1LG2*	eQTLgen	0.01	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	PLGRKT	eQTLgen	0.99	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	PPAPDC2	eQTLgen	-	-	-	0.00	0.13	0.23	0.00	0.00	0.00
rs13164856	*AC116366.6*	eQTLgen	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00
rs2271194	*CD63*	eQTLgen	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	DNAJC14	eQTLgen	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.01
	ERBB3	eQTLgen	0.11	1.00	1.00	0.03	0.00	0.00	0.11	1.00	1.00
	METTL7B	eQTLgen	-	-	-	0.00	0.14	0.03	0.00	0.00	0.00
	MYL6	eQTLgen	0.00	0.04	0.27	0.00	0.00	0.00	0.00	0.00	0.00
	ORMDL2	eQTLgen	-	-	-	0.00	0.00	0.00	0.00	0.00	0.02
	RDH5	eQTLgen	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	RPS26	eQTLgen	0.04	0	1.00	0.04	0.00	0.00	0.04	0.00	1.00
	SARNP	eQTLgen	0.05	0.16	0.76	0.01	0.62	0.00	0.04	0.16	1.00
	WIBG	eQTLgen	0.00	0.00	0.78	0.00	0.00	0.00	0.00	0.00	0.66
rs7563201	*AC010883.5*	eQTLgen	0.14	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	PLEKHH2	eQTLgen	0.00	0.00	0.89	0.00	0.00	0.00	0.00	0.00	0.49
	THADA	eQTLgen	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
	ZFP36L2	eQTLgen	0.3	1.00	1.00	0.00	0.31	0.08	0.16	1.00	1.00
rs7864171	*ZNF169*	eQTLgen	0.00	0.93	0.86	0.00	0.00	0.00	0.00	0.00	0.00
rs804279	*NEIL2*	eQTLgen	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
rs853854	*HCK*	eQTLgen	-	-	-	0.00	0.02	0.06	0.00	0.00	0.00
	RP5-1085F17.3	eQTLgen	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03

"-" signifies that the analyses could not be performed due to missing data to perform the analysis. N gives the setting for the maximum number of causal SNPs in the analysis.

4.3.3. Inadvertently selecting genetic variants based on P-value may yield false positive results

As evaluations of eCAVIAR suggested that the results may not be trustworthy, another colocalisation software called Coloc was used instead (Giambartolomei *et al.*, 2014; Hormozdiari *et al.*, 2016). When first conducting the Coloc analyses, the dataset based on the 10,000 SNPs most robustly associated with PCOS was used, where the summary statistics were based on analyses including the 23andMe cohort but with a limited number of SNPs. However, upon closer inspection of the results, it became clear that using this dataset potentially could give false positive results.

One example of a potential false positive is the Coloc analyses of *CAPS2* expression in transverse colon and PCOS risk (using 2 Mb region sizes). There was nominal evidence of colocalisation (PP=0.56) when using the top 10,000 SNPs dataset, but no evidence (PP=0.00) when using the main PCOS dataset. Visual inspection of the region using the main PCOS dataset suggested that SNPs in moderate LD with the top PCOS SNP are the SNPs most robustly associated with *CAPS2* expression, thus suggesting two separate causal variants for the two traits (the peak of dark and light blue SNPs in Fig 4.5). However, as the SNPs most robustly associated with *CAPS2* expression were only weakly associated with PCOS risk, these were excluded in the analyses using the top 10,000 SNPs dataset (Fig 4.6). In summary, pre-selecting SNPs based on P-value may inadvertently exclude SNPs that provide evidence against colocalisation. Therefore, all analyses using the top 10,000 SNPs dataset where disregarded. Instead, a combined version of the two publicly available PCOS datasets were used in the main Coloc analyses, including all SNPs and with effect estimates preferentially from the analyses including the 23andMe cohort.

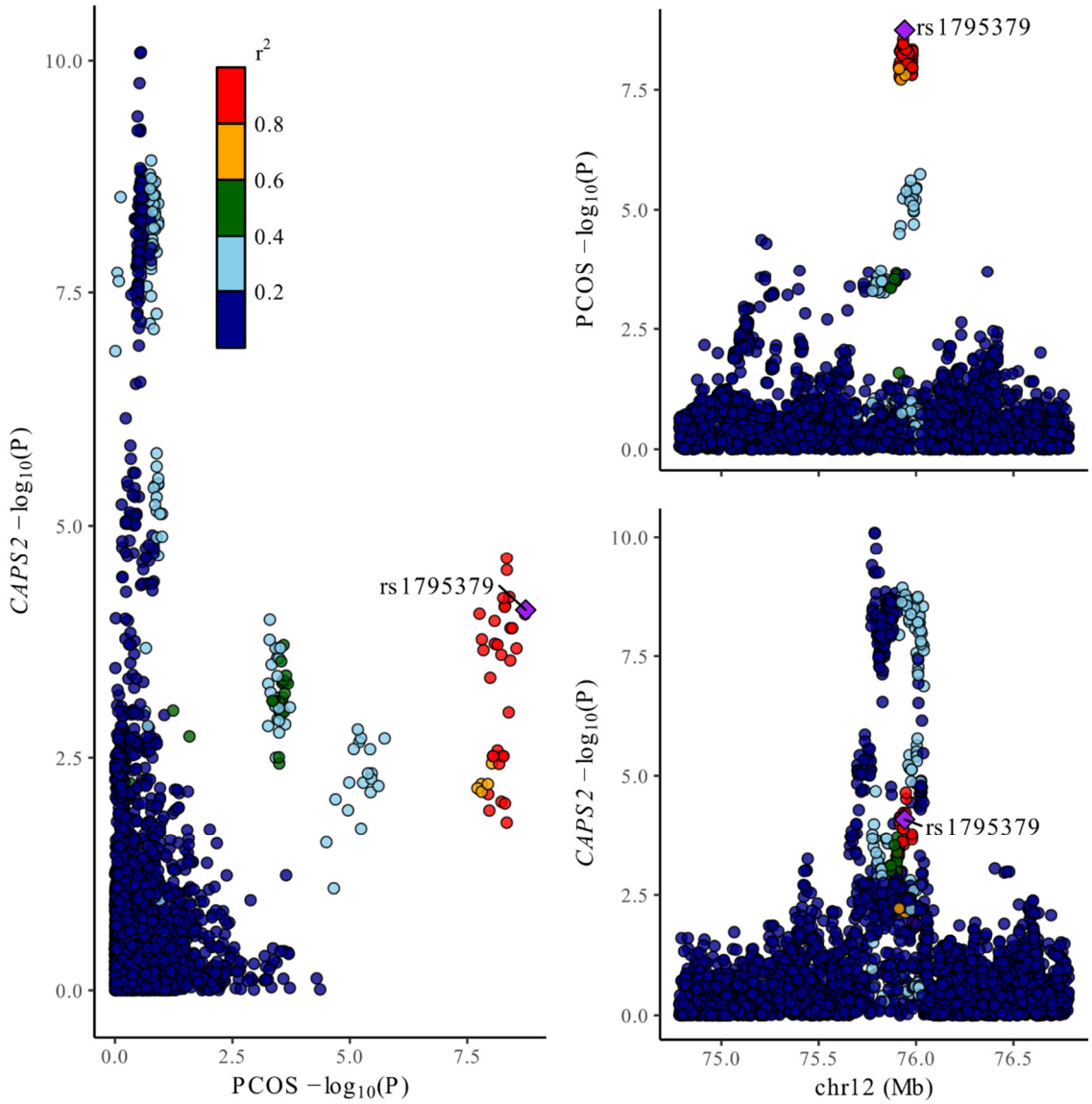


Figure 4.5. The SNP associations for PCOS risk (using the main PCOS dataset) and *CAPS2* expression in transverse colon (GTEx) at the rs1795379 risk locus. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r^2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

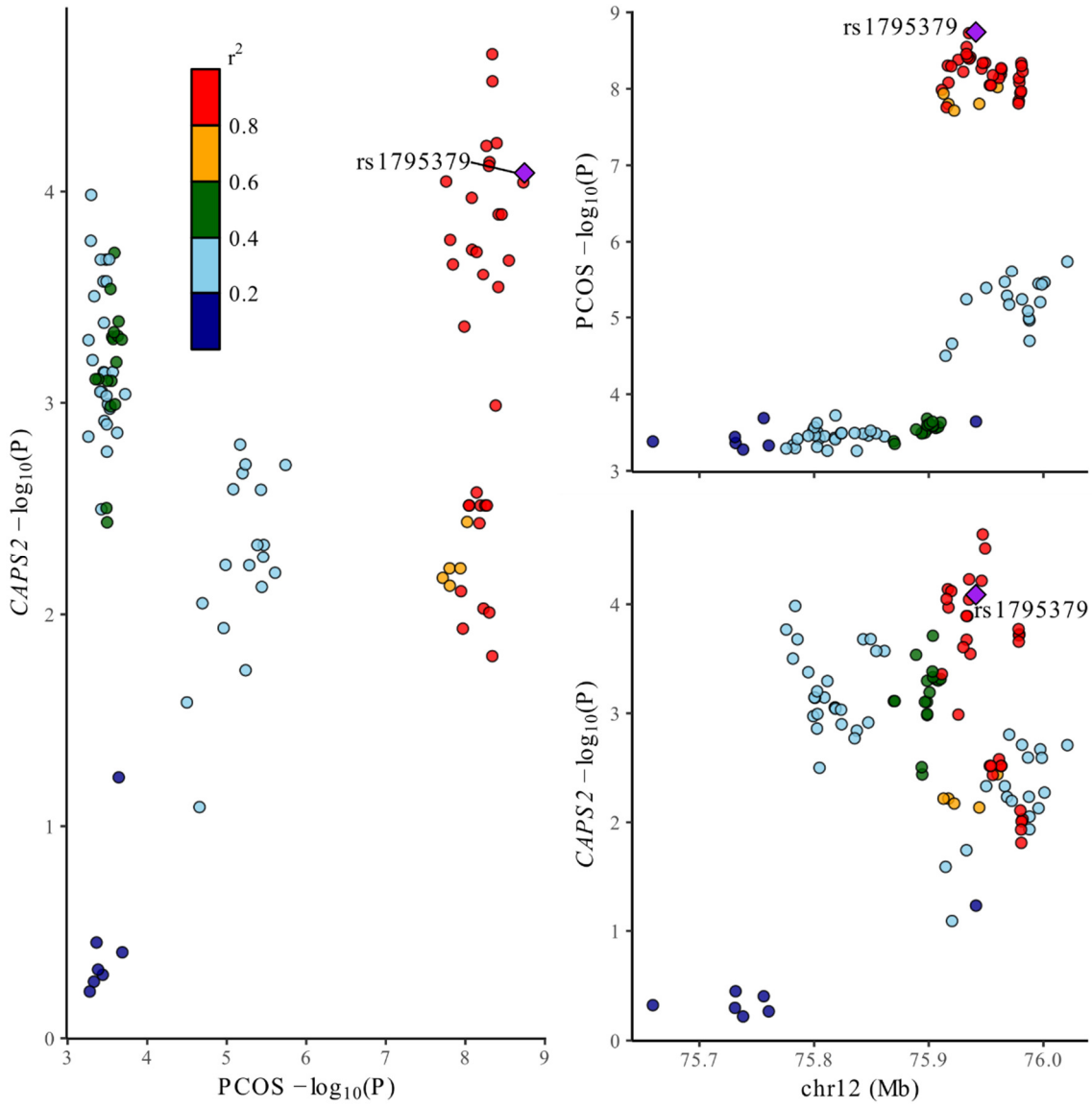


Figure 4.6. The SNP associations for PCOS risk (using the top 10,000 SNPs dataset) and *CAPS2* expression in transverse colon (GTEx) at the rs1795379 risk locus. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r^2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

4.3.4. Main Coloc results

4.3.4.1. Colocalisation results for polycystic ovary syndrome

Six genes and one protein had evidence (PP≥0.75) of colocalising with risk of PCOS, including FSH (PP=0.76), *ERBB3* (PP=0.94), *IKZF4* (PP=0.90), *RPS26* (PP=0.91), *SUOX* (PP=0.92), *ZFP36L2* (PP=0.93), and *C8orf49* (PP=0.87) (Fig 4.7 and Appendix Table 4.4). A further four

genes had nominal evidence (PP>0.50) of colocalisation, including *RAD50* (PP=0.53), *GDF11* (PP=0.53), *C9orf3* (PP=0.60), and *NEIL2* (PP=0.73). Some of these genes and proteins – such as FSH, *RAD50*, and *ZFP36L2* – had evidence of colocalisation in single tissue, whereas other genes were colocalising in a majority of the tested tissues (especially *RPS26* and *SUOX*). At four of the risk loci there was only evidence in favour of colocalisation for a single gene or protein, but a total of five genes were colocalising with disease risk at the rs2271194 risk locus. Visual inspection of the associations between the genetic variants and both PCOS and the quantitative traits generally supported the findings (Fig 4.8-4.10 and Appendix Fig 4.4-4.11).

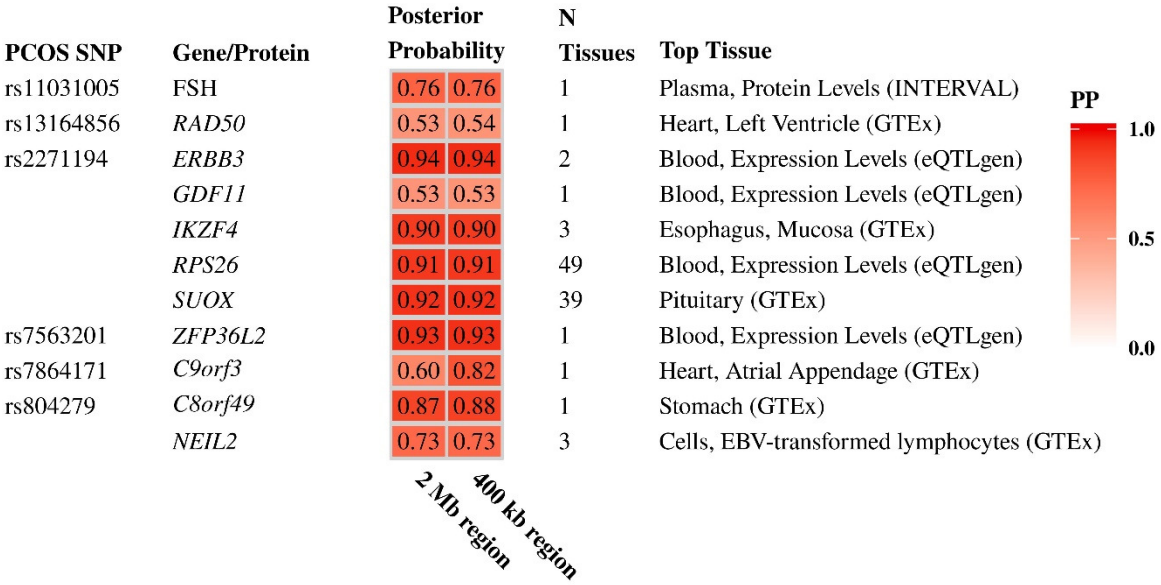


Figure 4.7. Colocalisation results for the genes and proteins with any evidence of colocalising with risk of PCOS. N tissues is the total number of tissue types and datasets that the gene/protein was colocalising in PP>0.50, and top tissue the tissue type with the highest PP. Figure adapted (changed layout and added information) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

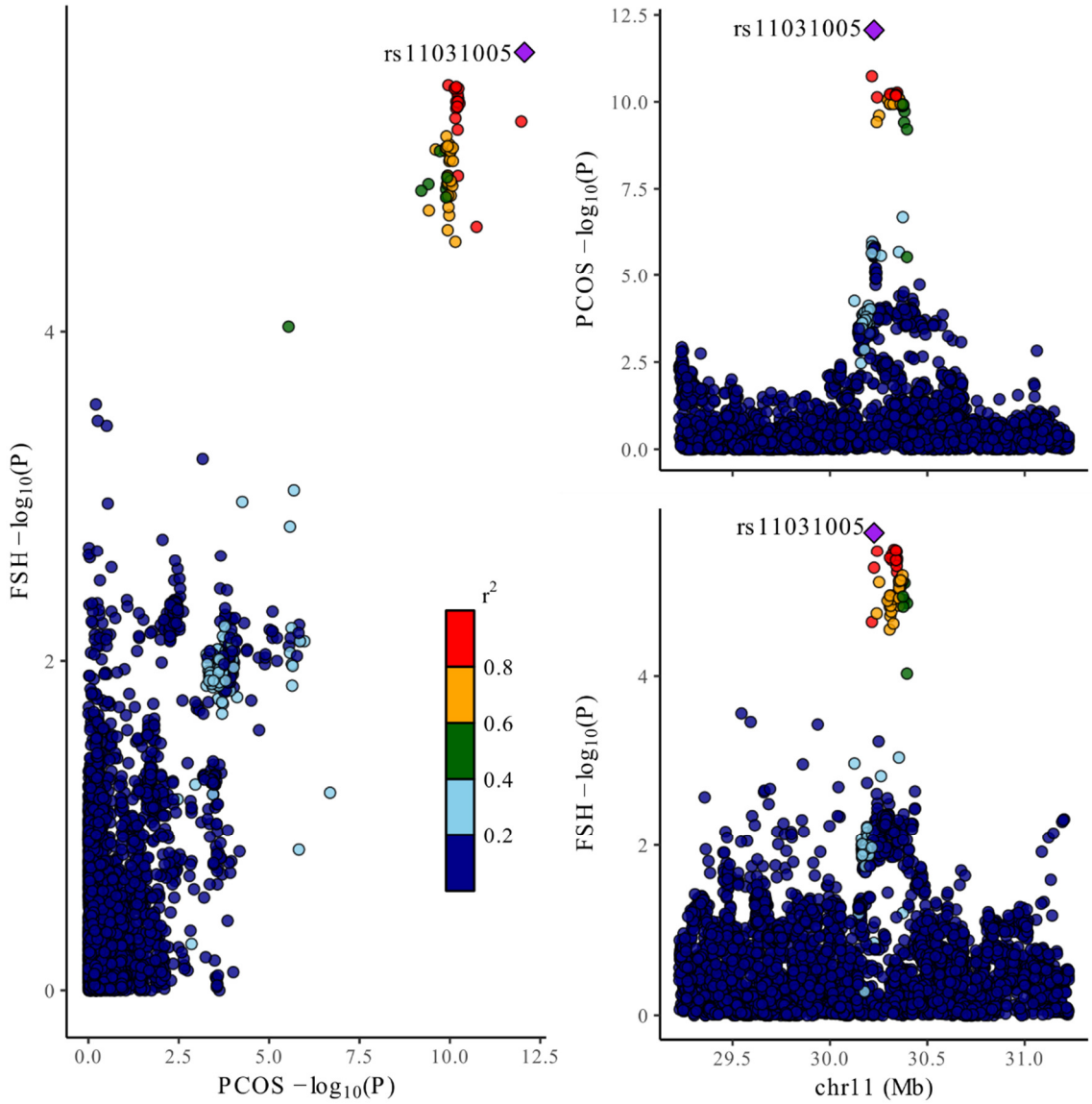


Figure 4.8. The SNP associations for PCOS risk (using the main dataset) and FSH levels in blood (INTERVAL) at the rs11031005 risk locus. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r^2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

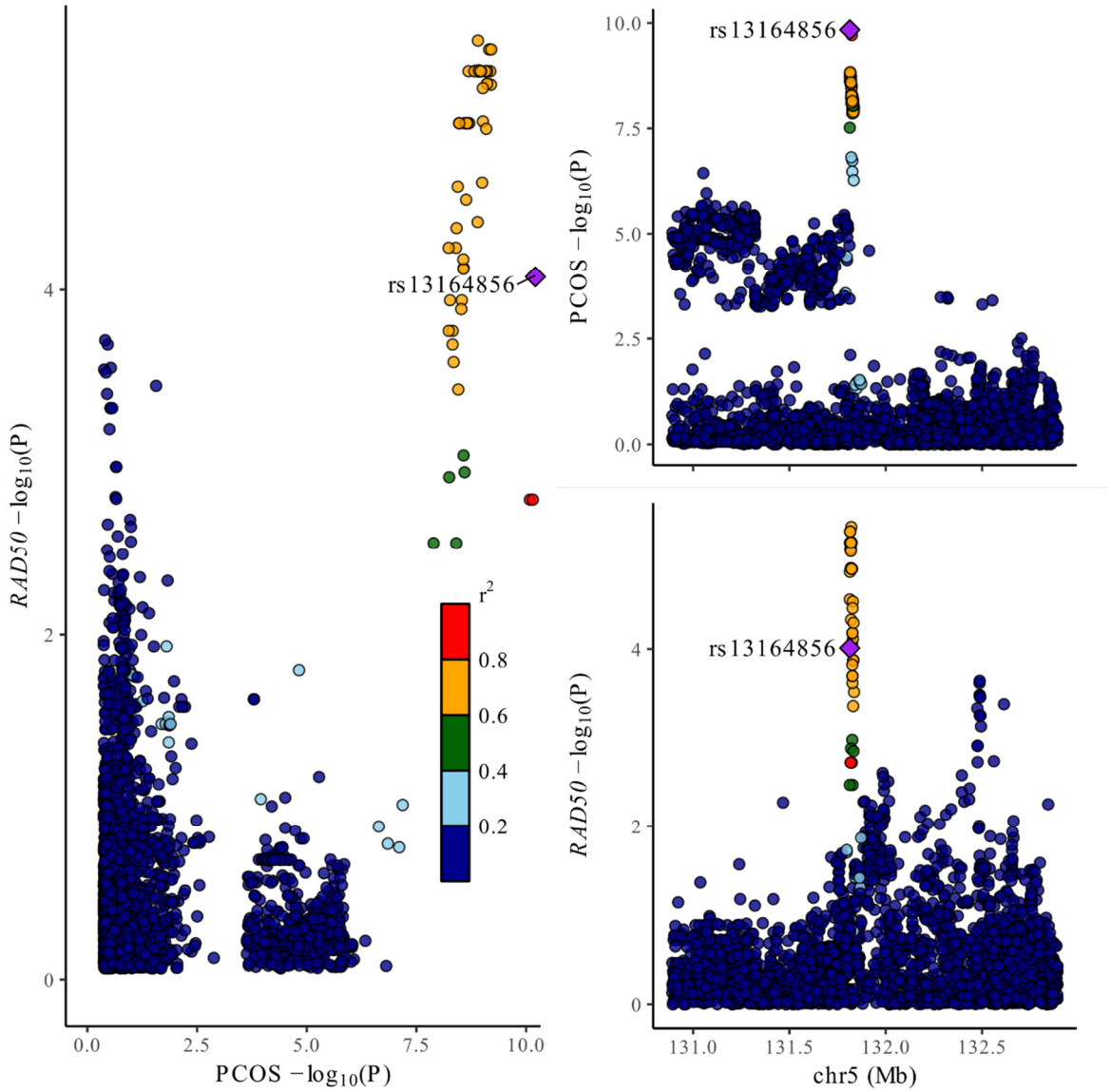


Figure 4.9. The SNP associations for PCOS risk (using the main dataset) and *RAD50* expression in the left ventricle of the heart (GTEx) at the rs13164856 risk locus. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r^2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

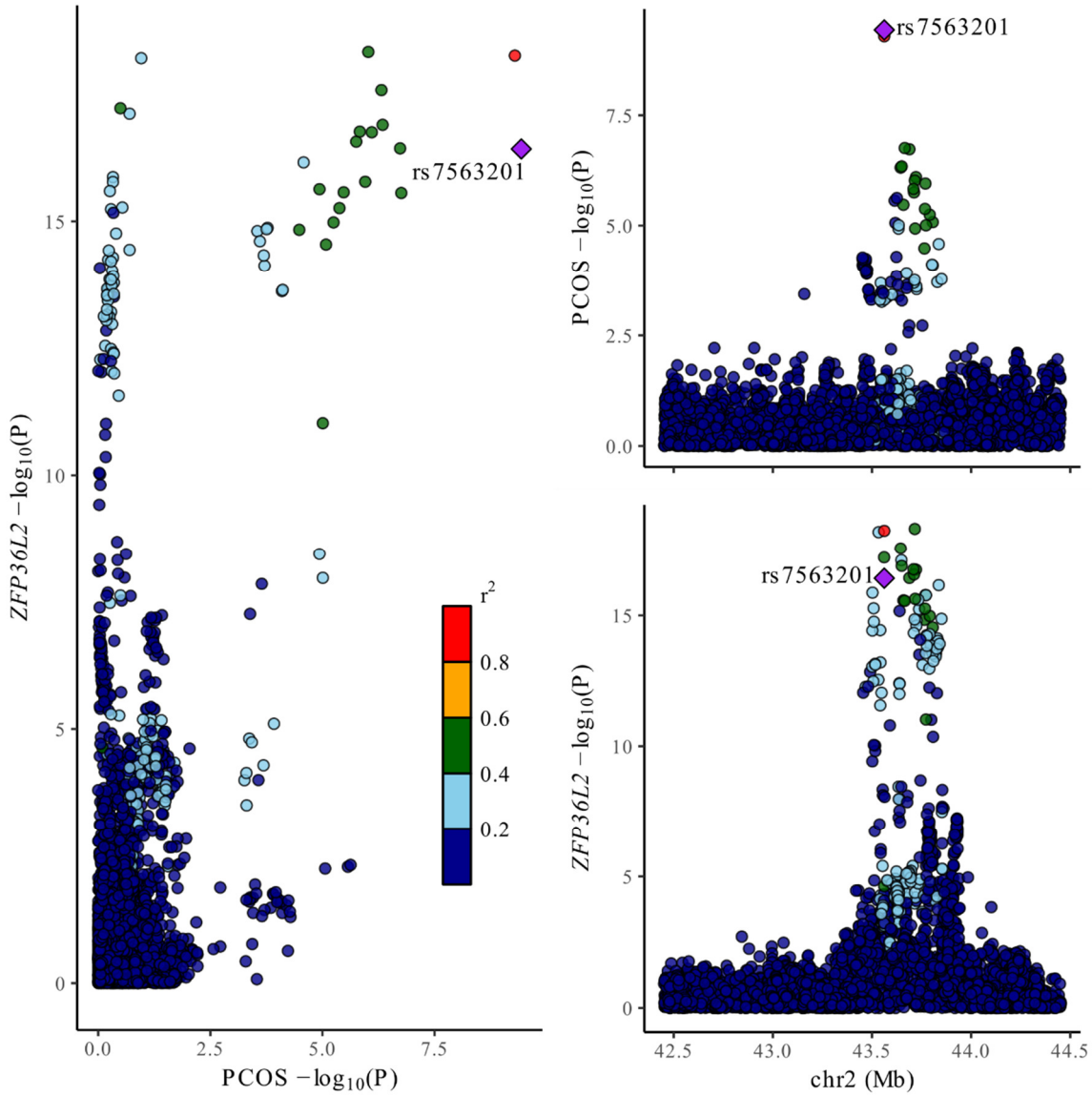


Figure 4.10. The SNP associations for PCOS risk (using the main dataset) and *ZFP36L2* expression in blood (eQTLgen) at the rs7563201 risk locus. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r^2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

4.3.4.2. No genes or proteins were highlighted as potential mediators for recurrent miscarriage

No gene or protein had any evidence of colocalisation (PP>0.50) with recurrent miscarriage using any region size, with or without conditioning on top hits. No additional analyses were therefore conducted for miscarriage.

4.3.5. Sensitivity analyses

4.3.5.1. Sensitivity analyses using a polycystic ovary syndrome dataset with approximately equal sample sizes for all genetic variants

Coloc uses summary statistics to compute PPs (Giambartolomei *et al.*, 2014). As summary statistics are affected by sample size, I was uncertain if using the main PCOS dataset – where different SNPs would have had effect estimates computed using very different sample sizes – would affect my results. Therefore, analyses were also conducted using the PCOS dataset where the 23andMe cohort had been excluded, to ensure that all SNPs had a similar sample size.

In general, Coloc analyses using this smaller PCOS dataset had both much lower power and lower PP (Fig 4.11 and Appendix Table 4.4-4.5). Still, five genes/proteins had a PP>0.50 using this PCOS datasets in at least one tissue and region size, including FSH, *GDF11, NEIL2*, *RPS26*, and *SUOX*.

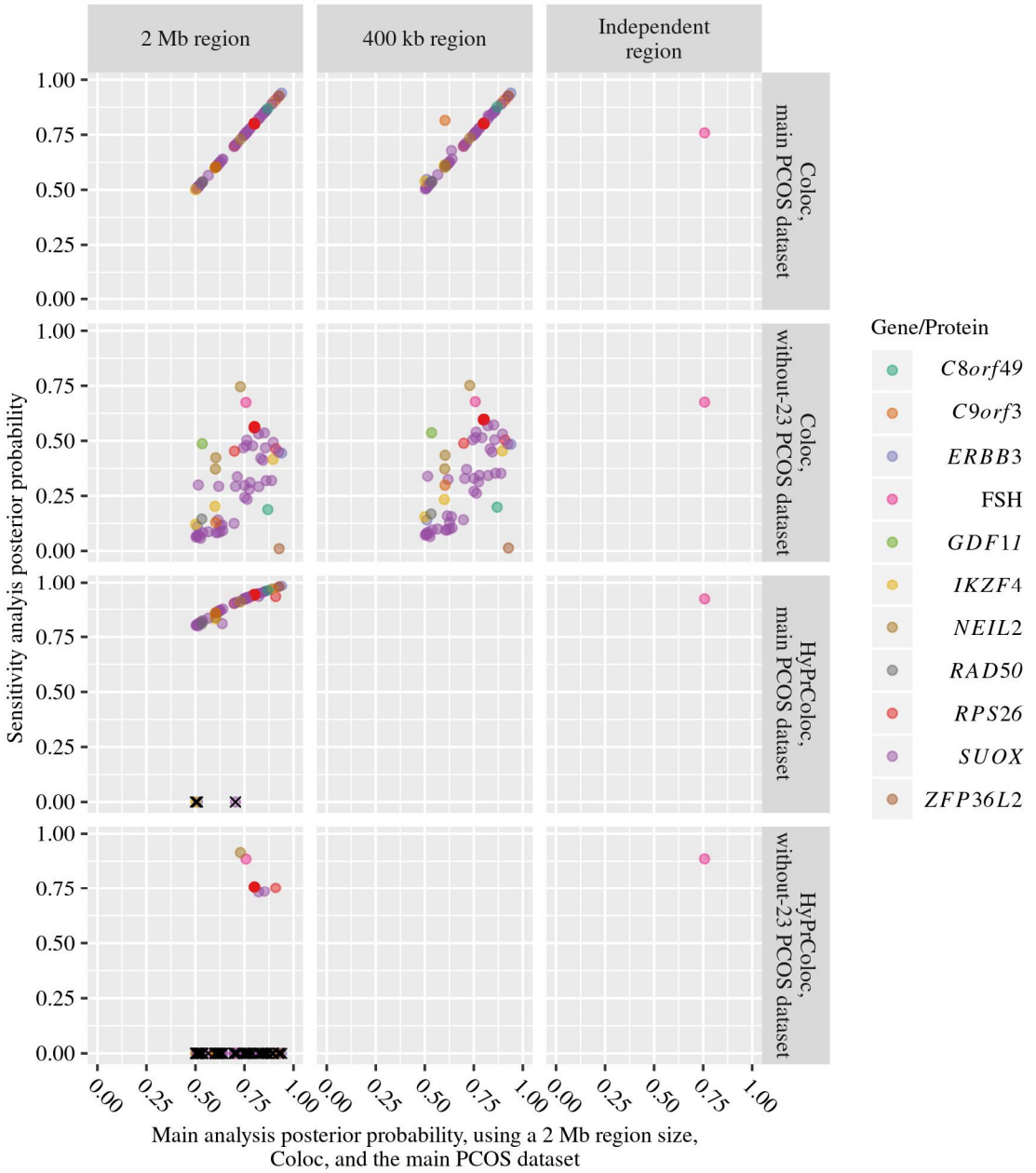


Figure 4.11. PPs from the sensitivity analyses of using different region sizes, PCOS datasets, and colocalisation software for the gene/protein-tissue combinations identified as colocalising with PCOS in the main analysis. The top left plot is a reference plot with the main results on both axes. The gene/protein-tissue combinations that had no evidence of colocalisation in the HyPrColoc analysis had no exact PP provided by the software; these analyses were manually assigned a PP of 0 and marked with an "X" for full disclosure. "Without-23" denotes the PCOS dataset where the 23andMe cohort had been excluded. Figure from Censin *et al.* (2021).

4.3.5.2. Region sizes

A simulation study has indicated that the PP of colocalisation is affected by the number of SNPs in the analysis (Wallace, 2020). Therefore, Coloc analyses were also conducted using a 400 kb region

size in all e/pQTL datasets and approximately independent LD regions in the INTERVAL dataset (Berisa and Pickrell, 2016; Sun *et al.*, 2018). These sensitivity analyses gave strong evidence in favour of the colocalising genes/proteins identified in the main Coloc analyses (Fig 4.11). All gene/protein combinations in a tissue with evidence of colocalisation (PP>0.50) identified in the main Coloc analyses had at least nominal (PP>0.50) evidence in the analyses using the other region sizes (using the main PCOS dataset).

4.3.5.3. HyPrColoc

In general, the results using HyPrColoc strongly supported the main results. All gene/protein and tissue combinations identified in the main analysis had evidence of colocalisation using HyPrColoc in all cases except three (Appendix Table 4.6). The HyPrColoc results for *SUOX* in blood (using the eQTLgen dataset), *IKZF4* in suprapubic skin, and *ERBB3* in spinal cord did not have evidence of colocalisation using HyPrColoc. However, there was still evidence of colocalisation with PCOS for these genes in other tissue types or datasets, still strengthening evidence of the genes – if not the tissues – being involved in PCOS pathophysiology.

4.3.5.4. Interaction-Coloc

Only four of the 11 genes and proteins identified in the main Coloc analyses had any interacting genes/proteins that Interaction-Coloc could be performed on (FSH, *RAD50*, *ERBB3*, and *C9orf3*). Thus, the possibility to detect additional evidence using Interaction-Coloc varied for the originally identified genes/proteins.

Still, the Interaction-Coloc analyses highlighted three genes linked to any of the ones originally identified (Fig 4.12). Of these, one gene – *RNF41* – is in the same genomic region as its originally identified gene *ERBB3*. Therefore, the Interaction-Coloc analyses did not provide "independent" evidence for *ERBB3* being a mediator of the reproductive disease. However, the Interaction-Coloc analyses also highlighted two other genes – *FSHR* (linked to FSH) and *UIMC1* (linked to *RAD50)*. Visual inspection of the *UIMC1* region supported colocalisation (Fig 4.13). For *FSHR*, visual

inspection supported colocalisation between *FSHR* gene expression and disease risk, but additionally suggested the presence of two independent peaks associating with the disease (Appendix Fig 4.12).

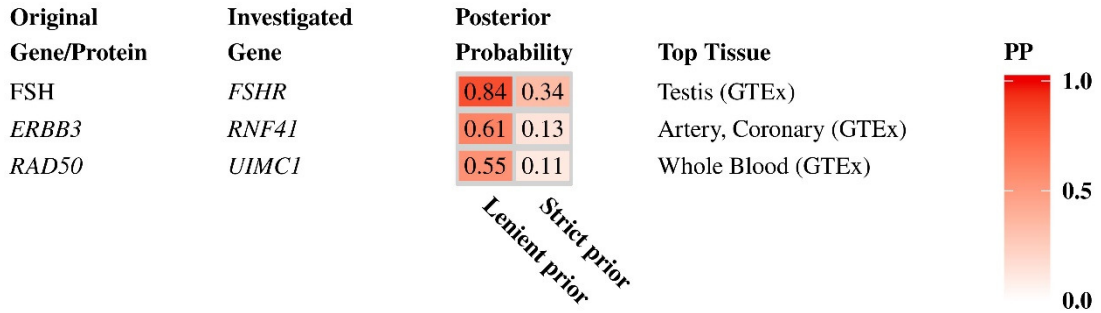


Figure 4.12. Interaction-Coloc results for the genes and proteins with any evidence of colocalising with risk of PCOS. The originally identified genes and proteins were identified in the main colocalisation analyses of genes and proteins in the regions around SNPs strongly associated ($P<5\times10^{-8}$) with PCOS. The investigated genes had evidence of being "linked" to the originally identified genes/proteins in protein-protein interaction data (Fabregat *et al.*, 2018). Colocalisation analysis was then performed to assess the evidence of a shared causal variant between the investigated genes and risk of PCOS using Coloc, 2 Mb region sizes, and two different prior settings (default, more lenient, Coloc p12 prior of $p12=1\times10^{-5}$ and stricter Coloc p12 prior of $p12=1\times10^{-6}$). Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

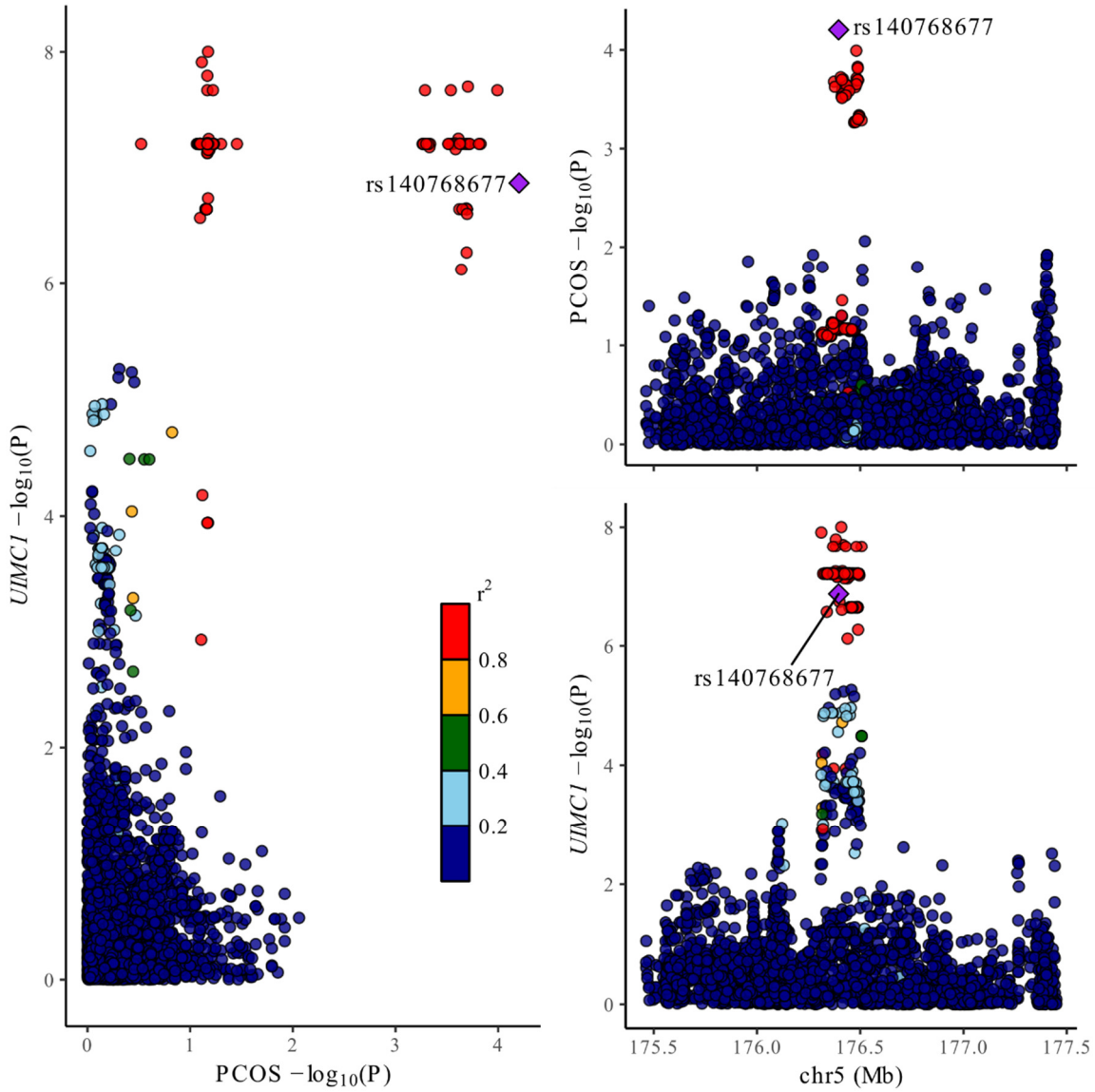


Figure 4.13. The SNP associations for PCOS risk (using the main dataset) and *UIMC1* (protein interacting with *RAD50*) expression in blood (GTEx). The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r^2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

The Interaction-Coloc results for *FSHR* using effect estimates conditioned on each of the two top index variants (rs4953650 and rs2349415) indicated that the colocalisation was driven by the rs4953650 peak (Appendix Fig 4.13-4.14). The Coloc results for *FSHR* using effect estimates conditioned on rs2349415 had a PP of 0.78, whereas the PP was 0.02 when using effect estimates conditioned on rs4953650.

4.3.6. Regulatory functions and associations with other traits

The results suggested that FSH may be mediating the disease risk at the rs11031005 locus. The risk allele for PCOS at rs11031005 (the C-allele) was associated with lower FSH levels (β -0.166 SD, 95% CI -0.235 to -0.097, $P=2.0\times10^{-6}$) (Day *et al.*, 2018; Sun *et al.*, 2018). The SNP was also associated with several other traits related to female hormonal regulation in the phenome-wide association study (Appendix Table 4.7). The two traits that were most robustly associated with rs11031005 was "length of menstrual cycle" ($P=1.2\times10^{-42}$) as well as "age at menopause" ($P=1.4\times10^{-15}$) (Carvalho-Silva *et al.*, 2019; Neale lab, 2020).

At the rs2271194 locus, there was evidence of colocalisation for several different genes (Fig 4.7). The data from the phenome-wide association study similarly indicated that the locus is very pleiotropic; the variant was associated with many different types of traits, for example hematologic, socioeconomic, and anthropometric traits (Appendix Table 4.8). *In-silico* look-up suggested that the SNP and/or its proxies (r^2 threshold of 0.8 in Europeans) have a regulatory function in several different cell types, including cervical cancer and hepatocellular cancer cells, immune cells, and brain cells (ENCODE Project Consortium, 2011; Ward and Kellis, 2012; Roadmap Epigenomics Consortium *et al.*, 2015). For the phenome-wide association look-up for the other loci, see Appendix Tables 4.9-4.11 (no trait was associated with rs804279).

4.4. Discussion and Conclusion

My third research aim was to establish if obesity has an effect on PCOS and recurrent miscarriage, and to attempt to identify potential mediating genes for these reproductive diseases. In this chapter, my work confirms a role of overall obesity in PCOS pathophysiology using MR, but found no evidence that obesity increases the risk of miscarriage. Colocalisation analyses highlighted several genes and proteins that may mediate the risk of PCOS, such as *ZFP36L2* and *RAD50,* and suggest that functional follow-up studies of PCOS-associated regions should focus on these candidates.

4.4.1. Mendelian randomisation analyses

4.4.1.1. Comparison with previous studies

The results gave strong evidence in favour of a causal role of overall obesity – measured as BMI – on risk of PCOS, in keeping with previous studies (Day *et al.*, 2018; Brower *et al.*, 2019). Furthermore, the results suggested a potential role of abdominal obesity – as measured by WHR – but these results were not strongly supported by sensitivity analyses or the analyses involving WHRadjBMI.

There was no evidence of obesity having an effect on risk of miscarriage or stillbirth, similarly to the MR study of BMI and sporadic miscarriage in one of the articles underlying this chapter (Laisk *et al.*, 2020). However, given the small sample sizes in GWASs of miscarriage and the many heterogeneous causes of the disease, the null results may be due to low power rather than an actual null effect (Larsen *et al.*, 2013; Zhou *et al.*, 2018; Laisk *et al.*, 2020).

4.4.1.2. Strengths and limitations of the Mendelian randomisation analyses

A major limitation in the MR analyses investigating the effect of obesity on miscarriage and stillbirth is the use of a different dataset with a slightly different phenotype than that of recurrent miscarriage (Day *et al.*, 2018; Zhou *et al.*, 2018). I considered simply meta-analysing the Z-scores of the obesity-related SNPs in the recurrent miscarriage dataset, but I have not seen this method to evaluate causality used in previous studies. I thus opted to use the larger dataset by Zhou *et al.* (2018), with the limitation of a broader case definition and likely more heterogenous set of cases.

4.4.2. Evaluation of eCAVIAR

Many colocalisation methods have been developed, and studies commonly use different settings even when using the same software (Giambartolomei *et al.*, 2014; Guo *et al.*, 2015, 2019; Hormozdiari *et al.*, 2016; Wen, Pique-Regi and Luca, 2017; Chun *et al.*, 2017; Aylward *et al.*, 2018; B. Liu, Calton, *et al.*, 2018; B. Liu, Pjanic, *et al.*, 2018; Franceschini *et al.*, 2018; Raulerson

et al., 2019; Çaliskan *et al.*, 2019; Foley *et al.*, 2021). I had first planned to use eCAVIAR for the colocalisation analyses, as it can account for allelic heterogeneity (Hormozdiari *et al.*, 2016). However, closer inspection of the results raised concerns that eCAVIAR may yield both false positive and false negative results. In addition, the results seemed highly dependent on the number of SNPs included in the analysis and the setting for maximum number of causal SNPs.

Only a few previous comparisons between colocalisation methods have been published (Chun *et al.*, 2017; Wen, Pique-Regi and Luca, 2017). The article describing a colocalisation software called *enloc* found that their results were more comparable to eCAVIAR than Coloc (Wen, Pique-Regi and Luca, 2017). However, another study by Chun *et al.* (2017) compared their colocalisation software JLIM to several other colocalisation methods, including Coloc, and found that Coloc generally performed quite well. Yet, my visual inspection of the genes and proteins denoted as colocalising in eCAVIAR versus Coloc suggested that Coloc gave more reliable results than eCAVIAR.

4.4.3. Selecting which genetic variants to include based on P-value may bias results

My results indicated that using a dataset where SNP inclusion was based on P-values can bias results, in essence because of the exclusion of SNPs that may provide evidence against colocalisation. In hindsight, this is not surprising. The risk of bias is clear when considering any two traits with separate lead SNPs in moderate LD with each other, which for example was illustrated in a previous article by Kristjansson *et al.* (2017).

4.4.4. Colocalisation analyses of polycystic ovary syndrome

4.4.4.1. Several of the colocalising genes are not the gene closest to the association peak

The colocalisation analyses identified eleven genes and proteins with any evidence of colocalisation – and a majority of these are not the gene closest to the PCOS index variant (Day *et al.*, 2018). For example, at the rs7563201 locus the closest gene is *THADA* and not *ZFP36L2,* and at the rs7864171 locus, the closest gene is *FANCC* and not *C8orf3* (Day *et al.*, 2018).

4.4.4.2. Colocalising genes and proteins emphasise a role of the hypothalamic-pituitary-gonadal axis in disease development

Much is still unknown about PCOS pathophysiology (Teede, Deeks and Moran, 2010). However, one contributing mechanism is likely perturbations in the hypothalamic-pituitary-gonadal axis that disrupts normal follicular processes such as ovulation (Azziz *et al.*, 2016). Of interest, several of the colocalising genes and proteins, including FSH, *RPS26, RAD50, ERBBR3*, and *ZFP36L2,* are connected to follicular development and/or the hypothalamic-pituitary-gonadal axis.

In normal menstrual cycles, the gonadotropins LH and FSH are secreted by the pituitary gland in response to oscillating levels of gonadotropin releasing hormone (Azziz *et al.*, 2016; Krishnan and Muthusami, 2017). However, women with PCOS have more rapid oscillations of gonadotropin-releasing hormone, which results in altered LH and FSH secretion patterns and a higher LH/FSH ratio (Yen, Vela and Rankin, 1970; Taylor *et al.*, 1997; Azziz *et al.*, 2016; Krishnan and Muthusami, 2017; Li *et al.*, 2019). These altered female hormonal levels may, in turn, contribute to the hyperandrogenism and perturbations of follicular development and ovulation observed in women with PCOS (Fauser and Van Heusden, 1997; Azziz *et al.*, 2016).

The results highlighted FSH as a potential effector gene at the rs11031005 risk locus. The phenome-wide association results also indicated that the locus is strongly associated with female reproductive traits. However, FSH share an alpha chain with several other hormones, including LH

(UniProt Consortium, 2019), and perturbations in the beta chain of FSH has been associated with higher LH levels in humans (Misgar *et al.*, 2019). Given the intricate relationship between the two hormones (Azziz *et al.*, 2016; Krishnan and Muthusami, 2017), it may be that the EA at the rs11031005 locus is contributing to PCOS risk by indirectly altering levels or secretion patterns of LH.

At the rs7563201 locus, the closest gene *THADA* has previously been suggested to be the effector gene (Pau *et al.*, 2017; Day *et al.*, 2018). However, the results gave no evidence in support of this theory, in spite of eQTL data for *THADA* being available in several tissues. Instead, the results suggested that *ZFP36L2* might be the disease-mediating gene at this locus. The protein product of *ZFP36L2* may be involved in regulating the levels of the LH receptor, and perturbation of the gene in mice disrupts oocyte development and ovulation (Ball *et al.*, 2014). Thus, there are several potential mechanisms through which *ZFP36L2* may mediate disease risk, and the gene warrants future functional follow-up.

Another promising gene candidate is *RAD50* at the rs13164856 locus. Even though the results only provided nominal evidence of colocalisation for *RAD50*, there was evidence of potential colocalisation for *UIMC1* in the Interaction-Coloc analyses. As *UIMC1* encodes a protein with links to the *RAD50* protein complex, in theory this should strengthen the likelihood of the *RAD50* being involved in the disease pathophysiology (Mailand *et al.*, 2007; Inagaki, Roset and Petrini, 2016; UniProt Consortium, 2019). Furthermore, the literature suggests that *RAD50* may be involved in oocyte development (Inagaki, Roset and Petrini, 2016), and disruptions of *RAD50* may reduce fertility in female mice (Roset *et al.*, 2014).

Five genes were colocalising with PCOS risk at the rs2271194 locus. Two of these – *RPS26* and *ERBB3* – have connections to the hypothalamic-pituitary-gonadal axis. Mouse studies have implicated *Rps26* in follicular development and ovarian response to gonadotropin simulation (X.-

M. Liu *et al.*, 2018). Gonadotropins may also upregulate *ERBB3* expression, and *ERBB3* may be involved in follicular development (Mukherjee and Roy, 2013; Chowdhury *et al.*, 2017). For the other genes, there is no clear potential mechanistic link to PCOS based on the literature. *IKZF4* seems to be involved in the immune response (Pan *et al.*, 2009; Gokhale *et al.*, 2019), but it is not clear how it would affect risk of PCOS. *GDF11* is an important gene during development (Wu *et al.*, 2003; Kim *et al.*, 2005; Liu, 2006; Luo *et al.*, 2019). While androgens may downregulate *GDF11* expression levels in mouse ovaries (Liu *et al.*, 2017), it is not otherwise clear how the gene would affect PCOS pathophysiology. One of the other colocalising genes at the rs2271194 locus, *SUOX,* is a mitochondrial sulfite oxidase (UniProt Consortium, 2019). *SUOX* is expressed in mice oocytes, but not much else is known about the gene and its function (Pan *et al.*, 2008; Biase, 2017).

NEIL2, a base excision repair gene, was colocalising with PCOS risk at the rs804279 loci (Chakraborty *et al.*, 2015). DNA damage repair is an integral part of normal oocyte development (Martin *et al.*, 2019), but knock-out of the gene in mice did not affect fertility (Chakraborty *et al.*, 2015; Rolseth *et al.*, 2017) and another study found undetectable levels of *NEIL2* in human oocytes (Jaroudi *et al.*, 2009). For the other gene colocalising at the rs804279 loci, *C8orf49*, as well as for the gene colocalising at the rs7864171 loci (*C9orf3*), very little is known about their potential function.

4.4.4.3. Possible extensions of Interaction-Coloc

A possible extension of Interaction-Coloc would be to perform colocalisation on all the genes and proteins that are in the same extended pathway as the originally identified genes and proteins, rather than only those with evidence of protein-protein interactions. However, a new software, Polygenic Priority Score (PoPS), was recently released that use such an integrative approach to identify causal genes in GWASs, where data from GWAS summary statistics, biological pathways, gene expression, and protein-protein interaction data are included (Weeks *et al.*, 2020). As PoPS is the product of a very large-scale effort to incorporate evidence from multiple sources, and hold promise for giving reliable results, I did not expand on the Interaction-Coloc method (Weeks *et al.*,

2020). Still, a possible future avenue would be to apply PoPS to the PCOS dataset and compare the results with the ones presented in this chapter.

4.4.4.4. Strengths and limitations of the colocalisation analyses

4.4.4.4.1. Strengths and limitations related to the methods

Great care was taken to assure the robustness of the results, including careful evaluation of the results using different colocalisation methods and visual inspection of the results (Giambartolomei *et al.*, 2014; Hormozdiari *et al.*, 2016; Liu, Gloudemans and Montgomery, 2018; Foley *et al.*, 2021). I also evaluated how using a dataset where SNPs had been pre-selected based on P-value affected the results, and found that this introduce bias. However, the use of a dataset where different SNPs have different sample sizes did not seem to bias the results. Still, this may only be true in cases where the sample size is determined *after* association testing in the full sample, and when it is only the more weakly associated SNPs that have summary statistics computed using fewer samples. More studies are needed to determine the implications of including SNPs with unequal sample sizes in other scenarios.

Coloc further assumes a single causal variant per locus (Giambartolomei *et al.*, 2014; Wallace, 2020). Yet, the software can give reliable results even in the case of allelic heterogeneity, unless the two signals have comparable strengths of association (Giambartolomei *et al.*, 2014; Wallace, 2020). Still, it is possible that allelic heterogeneity may have caused false negative results in my study.

Genes can share regulatory features, and several genes can colocalise with a trait at a locus (Bhatia and Kleinjan, 2014; Hormozdiari *et al.*, 2016; Krijger and De Laat, 2016; Yao *et al.*, 2017). In addition, the majority of eQTLs affect gene expression in multiple tissues (Ongen *et al.*, 2017; Wang *et al.*, 2019). Thus, several gene, tissue, and time-point combinations can be colocalising with a disease, even if only one is the true mediating gene, tissue and disease-relevant timepoint

combination. Indeed, the true causal gene, tissue, and disease-relevant timepoint combination may not even have been investigated, which may explain why eight PCOS loci did not colocalise with any genes or tissues. Another potential explanation is that the disease-risk is mediated through changed protein function rather than protein levels if the true causal SNP affects the coding sequence. Thus, even though colocalisation can highlight genes and proteins that may mediate risk of PCOS, the results should not be interpreted as definitive evidence.

In this chapter, I suggested a new extension of colocalisation methodology – Interaction-Coloc. Whereas this novel approach is promising and yielded reasonable results, the method has not been validated and the results should be interpreted with caution. As discussed together with my colleague Dr. Jonas Bovijn, a possible approach to further evaluate and apply the software could be to perform an Interaction-Coloc analysis on index variants identified in older summary statistics from larger GWAS consortia. In theory, one would then expect Interaction-Coloc to identify genomic regions that have been identified in later GWASs with larger sample sizes, which would render further credibility to the method.

4.4.4.4.2. Strengths and limitations related to the datasets

Genetic variants can be sex-differentially associated with, for example, expression levels (Oliva *et al.*, 2020). Thus, a major limitation of the study is the use of quantitative trait loci datasets that included both men and women, which potentially could bias the results if the quantitative trait loci associations differ between the sexes (GTEx Consortium, 2017; Sun *et al.*, 2018; Võsa *et al.*, 2018). Another potential issue relating to the INTERVAL dataset is the possibility of changed aptamer binding (Sun *et al.*, 2018). If SNPs change the coding sequence of a protein, they may cause changed aptamer binding and become false positive pQTLs (Sun *et al.*, 2018). Yet, FSH was the only colocalising protein, and the top PCOS SNP in that locus have previously been reported as associated with FSH levels using another protein quantification method, minimising the risk of bias from changed aptamer binding (Mbarek *et al.*, 2016),

Coloc analyses requires that the included datasets consists of participants of a similar ancestral background (Giambartolomei *et al.*, 2014). In this study, all datasets primarily consisted of participants of European descent, and the PCOS risk loci were restricted to those confirmed in a GWAS in Europeans (GTEx Consortium, 2017; Day *et al.*, 2018; Sun *et al.*, 2018; Võsa *et al.*, 2018; Kikas *et al.*, 2019; Laisk *et al.*, 2020). Yet, it is difficult to fully exclude potential bias from fine-scale population stratification or bias from the small proportion of non-European participants in, for example, GTEx (GTEx Consortium, 2017; Berg *et al.*, 2019; Haworth *et al.*, 2019). Finally, the use of datasets with predominantly European participants means that the findings may not be transferable to women of other ancestries.

4.4.5. Conclusion

My work in this chapter confirms a role of obesity in PCOS, but not in miscarriage. I highlight potential effector genes and proteins for PCOS, several of which have been previously linked to the hypothalamic-pituitary-gonadal axis and oocyte development. Some of the colocalising genes are not the closest gene to the risk loci and have received little attention for a potential role in PCOS pathophysiology in previous studies. In particular, my results emphasise that especially *RAD50, ZFP36L2, ERBB3, RPS26,* and FSH may have a role in PCOS pathophysiology and warrant future functional follow-up.

Chapter 5. Discussion and Conclusion

5.1. Discussion

5.1.1. A summary

Men and women are not biologically identical. This statement might seem obvious, given the numerous examples of sex differences in body morphology, hormonal levels, disease presentations, drug effects, and other areas (The Steering Committee of the Physicians' Health Study Research, 1988; Schreiner *et al.*, 1996; United States General Accounting Office, 2001a; Marshall, 2005; Ridker *et al.*, 2005; Demerath *et al.*, 2007; DeVon *et al.*, 2008; Karakas *et al.*, 2018; Khan *et al.*, 2018; Ruth *et al.*, 2020). Yet, historically, there has been an underappreciation of the importance of sex and sex differences in research (Meinert *et al.*, 2000; United States General Accounting Office, 2001b; Beery and Zucker, 2011; Geller *et al.*, 2011, 2018; Turcot *et al.*, 2018). I therefore wanted to focus on sex-specific effects and sex differences to try and improve healthcare for both men and women in my thesis. More specifically, I decided to explore sex-specific effects in the causes and consequences of obesity, as obesity is responsible for millions of deaths every year (NCD Risk Factor Collaboration (NCD-RisC), 2016b; GBD 2015 Obesity Collaborators *et al.*, 2017). To this end, I had three overarching research aims:

I. Determine the effect of obesity in leading mortality causes in men versus women (Chapter 2).
II. Discover potential sex-specific protein biomarkers and drug targets for obesity and leading cardiometabolic diseases (Chapter 3).
III. Establish the effect of obesity on two obesity-associated female reproductive diseases and identify potential disease-mediating genes (Chapter 3).

The first two research chapters in my thesis investigate overarching questions regarding sex-specific causes and consequences of obesity. In Chapter 2, I looked at the sex-specific consequences of obesity on the leading causes of death from non-communicable diseases (The World Health Organization, 2018b). Using MR, I find that obesity has wide-spread effects on the majority of leading causes of death in both men and women (Davey Smith and Ebrahim, 2003). I also establish that increased obesity may sex-differentially impact on risk of disease, including on risks of T2D, COPD, and renal failure.

The work presented in Chapter 3 aimed to contribute towards a solution of the obesity pandemic by identifying proteins that contribute to obesity and cardiometabolic disease in men and women, and to explore how any effects differ between the sexes. My results in this chapter suggest that genetic variants can be sex-differentially associated with protein levels, and that proteins may have different effects on obesity in men versus women. I identify several proteins that may be valuable as sex-specific biomarkers or drug targets in obesity. For example, higher levels of kynureninase may protect against obesity in women, but increase obesity in men. However, there was evidence of the underlying aptamer data in Chapter 3 being affected by poor aptamer specificity, which could cause false positive and negative results.

Finally, in Chapter 4, I focused on two common obesity-related reproductive diseases in women with considerable consequences on patient health; PCOS and miscarriage (Fauser and Van Heusden, 1997; Meyer, McGrath and Teede, 2005; Trinder *et al.*, 2006; Cumming *et al.*, 2007; Yildiz, Knochenhauer and Azziz, 2008; Jones *et al.*, 2008; Moran *et al.*, 2010; Boots and Stephenson, 2011; Cantwell *et al.*, 2011; Joham *et al.*, 2015; Maino *et al.*, 2016; Weng *et al.*, 2018; Brower *et al.*, 2019; Cavalcante *et al.*, 2019; Damone *et al.*, 2019; Greenwood *et al.*, 2019). Research into these diseases have been hampered by small study sizes and, in certain cases, poor quality (Aleman *et al.*, 2005; Broder-Fingert *et al.*, 2009; Sivayoganathan *et al.*, 2011; Chen, Fu and Huang, 2016; Qiao *et al.*, 2016; Pereza *et al.*, 2017; Wahabi *et al.*, 2018; Lim *et al.*, 2019).

Thus, to contribute towards improved therapeutic options and increased understanding of these diseases, I first use MR to confirm obesity as a risk factor on PCOS, but could not confirm obesity as a risk factor for miscarriage (Davey Smith and Ebrahim, 2003). I then perform colocalisation to highlight potential disease-mediating genes and proteins for PCOS, including *RAD50*, *ZFP36L2*, *ERBB3*, *RPS26*, and FSH (Giambartolomei *et al.*, 2014). Unfortunately, I could not identify any potential effector genes for recurrent miscarriage.

Since I started the work included in this thesis, there seems to have been an increased appreciation of sex as a biological variable in the field. Several GWASs and MR studies that investigate sex-specific effects have been published (Lotta *et al.*, 2018; Ruth *et al.*, 2020; Lagou *et al.*, 2021). A recent, large-scale study in GTEx by Oliva *et al.* (2020) found that sex is an important determinant for gene regulation, but also that sex-differential gene regulation varies between tissues. Both Oliva *et al.* (2020) and another study by Porcu *et al.* (2020) also assessed the evidence for eQTLs having different effects in men versus women. In all, the studies by Oliva *et al.* (2020), Porcu *et al.* (2020), and others (Lotta *et al.*, 2018; Ruth *et al.*, 2020; Lagou *et al.*, 2021) highlight how my work is part of recent effort to determine sex-specific effects and sex-differences in gene regulation, phenotypes, and epidemiological relationships between traits and diseases.

5.1.2. Key limitations

5.1.2.1. Sex-differential participation bias and the impact of normalisation procedure

Although there is now a greater understanding of the importance of sex in research, it has also become clearer that there are methodological challenges in investigating sex differences using common methods in the genetic epidemiology field (Pirastu *et al.*, 2020). As discussed in Chapter 2, a study by Pirastu *et al.* (2020) highlighted the possibility of sex-comparisons being affected by sex-differential participation bias and unit of comparison. Of these two issues, the possibility of sex-differential participation bias is, at the moment, not possible to fully quantify or correct for (Pirastu *et al.*, 2020). Still, as discussed in Chapter 2, it seems unlikely that it is a major source of

bias in my work, as it is only under the most extreme simulation scenarios that sex-differential bias would cause apparent sex-differences on par with my results. In Chapter 3, I was not able to find any evidence of sex-differential participation bias in the INTERVAL data, but as I had a small sample size, I had low power to detect any bias. In summary, while more research is needed to fully understand to what degree sex-differential participation bias might affect comparisons between the sexes, at present it seems unlikely to be substantial enough to be the primary cause to my results.

The second issue raised by Pirastu *et al.* (2020) is the impact of normalisation procedure, and consequently unit, on comparisons between men and women. I explored this question in detail in Chapter 2 and Chapter 3, and found that choice of normalisation procedure and unit have considerable effects on estimated sex differences. Yet, in my opinion, choice of unit is ultimately dependent on what type of sex differences one is interested in. As an example, we can consider height, as men tend to be taller than women on average (NCD Risk Factor Collaboration (NCD-RisC), 2016a). Are we interested in understanding if genetic variants are associated with different increases of height in "absolute" units (such as centimetres), or different increases of height in "relative" units (such as sex-specific SD units), in men versus women? In my opinion, sex-comparisons that account for the overarching differences in a trait between men and women would generally be of more interest.

5.1.2.2. Limitations inherent to Mendelian randomisation

MR methods were used in all three research chapters (Davey Smith and Ebrahim, 2003). Yet, MR is based on several assumptions, of which not all are directly verifiable (Greenland, 2000; Martens et al., 2006; Burgess et al., 2015).

First of all, the genetic instruments must be associated with the exposure (Greenland, 2000; Martens et al., 2006; Burgess et al., 2015). In addition, genetic instruments that are only weakly associated with the exposure may lead to weak instrument bias (Staiger and Stock, 1997; Burgess,

Davies and Thompson, 2016; Burgess, Dudbridge and Thompson, 2016). Throughout my thesis, I used instruments that were strongly associated with the exposure. Yet, even if genetic variants surpass a strict multiple testing threshold, they may represent false positive associations due to chance or have inflated effect estimates due to winner's curse (Taylor *et al.*, 2014). Furthermore, for the MR analyses performed in Chapter 3, changed aptamer binding could result in coding SNPs becoming false positive cis-pQTLs, which also could violate the criterium of the instrument being associated with the exposure (in this case, protein *levels*) (Greenland, 2000; Martens *et al.*, 2006; Burgess *et al.*, 2015; Sun *et al.*, 2018).

Another key MR assumption is that the genetic instruments are not associated with any potential confounders of the exposure-outcome relationship, which is known as horizontal pleiotropy (Greenland, 2000; Martens *et al.*, 2006; Burgess *et al.*, 2015, 2017). The challenge of this criterium is that it is not fully testable (Burgess *et al.*, 2017). It is often assumed that larger number of instruments decreases the risk of directional bias in the MR estimate, under the assumption that the pleiotropic effects from individual instruments balance each other out (Burgess *et al.*, 2017). Yet, a recent study found that 48% of MR estimates had evidence of horizontal pleiotropy (Verbanck *et al.*, 2018). In order to minimise the risk of pleiotropy affecting the results, I conducted analyses more robust to such bias where possible (Bowden, Davey Smith and Burgess, 2015; Bowden *et al.*, 2016). For Chapter 3, I additionally tested the associations between each protein instrument against all other proteins in the dataset. Still, a possible extension for all my research chapters would be to reconduct the MR analyses using the newly published CAUSE method, which seems to lessen the risk of bias from horizontal pleiotropy while retaining a high power to identify causal effects (Morrison *et al.*, 2020). Importantly, these MR limitations could potentially affect my results on two levels. Firstly, by affecting the MR estimates, but they could also bias the MR estimates *sex-differentially*, and thus lead to false positive sex differences.

Finally, an intervention on a risk factor may not yield the same effect as that estimated from an MR analysis, for example due to different durations of the exposure and biological buffering against genetic effects (Waddington, 1942; Lawlor *et al.*, 2008; Burgess, Davey Smith, *et al.*, 2020). In addition, the effect of an exposure on an outcome may change over the course of life (van den Brandt *et al.*, 2000; Guo *et al.*, 2016; Al-Ajmi *et al.*, 2018; Shu *et al.*, 2018). For example, there is some evidence that the effect of BMI on breast cancer may be directionally opposite for pre- versus post-menopausal breast cancer (van den Brandt *et al.*, 2000; Guo *et al.*, 2016; Al-Ajmi *et al.*, 2018; Shu *et al.*, 2018).

5.1.2.3. Non-specific aptamer binding

For Chapter 3, one of the key limitations is the poor agreement between aptamers measuring the same protein, likely caused by non-specific aptamer binding. The results in this chapter should therefore be seen as suggestive, and need to be replicated in another cohort using another protein quantification method. Still, the fact that a proportion of the pQTLs did replicate supports the existence of true sex-differential pQTLs.

5.1.2.4. Eurocentric bias

Individuals of European ancestry are vastly overrepresented in GWASs (Mills and Rahal, 2019). This is a major problem for several reasons. Firstly, it is possible that epidemiologic relationships between risk factors and diseases differ between ancestries. For example, the risk of T2D increases at different obesity levels in different ancestries, which has prompted the International Diabetes Federation to recommend ancestry-specific WHR cut-off values for being at increased risk of diabetes (International Diabetes Federation, 2006; Frank *et al.*, 2013; Babai *et al.*, 2016). In addition, population differences in LD patterns and genetic effects of SNPs can affect the generalisability of results from studies in Europeans to those of other ancestries (Shifman *et al.*, 2003; Mahajan *et al.*, 2014; Berisa and Pickrell, 2016; Márquez-Luna *et al.*, 2017; Curtis, 2018; Duncan *et al.*, 2018; Grinde *et al.*, 2019; Cavazos and Witte, 2021). Therefore, the Eurocentric bias poses an ethical problem, as scientific advancements using GWAS data from Europeans are likely

to benefit individuals differently depending on ancestry. Finally, the Eurocentric bias also hampers research. For example, fine-mapping efforts to identify the causal variants at a trait-associated loci are improved by the inclusion of trans-ethnic data (Mahajan *et al.*, 2014; Mägi *et al.*, 2017). Thus, a major limitation of my work – and the genetic epidemiology field in general – is that the analyses were conducted almost exclusively in individuals of European ancestry.

5.1.3. Directions for future research

5.1.3.1. In-depth characterisation of sex-specific effects

My work presents many examples of sex differences in the causes and consequences of obesity and highlight genetic variants with sex-differential effects on protein levels. Yet, for the vast majority of my results, the exact mechanisms behind the observed sex differences are unknown. For example, it is unclear to what extent the sex differences are due to relative differences between body morphology and fat distribution in men versus women (Schreiner *et al.*, 1996; Demerath *et al.*, 2007), how transcription factors are regulated by sex hormones (Oliva *et al.*, 2020), or how transcription factors target different gene regulatory networks in men versus women (Lopes-Ramos *et al.*, 2020).

These, and other questions, need to be answered to fully understand the mechanisms behind the observed sex differences in the genetic associations and in the relationships between traits and diseases. To do this, two other key questions must be answered; what is the causal variant in each trait-associated locus, and which is the effector gene? To answer these questions will likely require integration of many different data sources, such as data on genetic associations, gene expression and protein levels, transcription factor binding in different tissues, biological pathways, and protein-protein interactions, ideally across different timepoints in the human lifespan. Still, there are continuous methodological advancements towards such an integrative approach, such as the newly developed Polygenic Priority Score (PoPS) (Pers *et al.*, 2015; Weeks *et al.*, 2020). These advancements hold promise for facilitating more in-depth understanding of how genetic variants

affect disease and, thereby, how effects differ between the sexes (Pers *et al.*, 2015; Weeks *et al.*, 2020).

5.1.3.2. Replication of sex-differential protein quantitative trait loci and protein-outcome relationships using another protein quantification method

The sensitivity analyses in Chapter 3 indicated poor agreement between pairs of aptamers targeting the same protein, suggesting that a proportion of the aptamers had poor specificity. Thus, the results presented in this chapter need be validated in a separate cohort using another large-scale protein quantification method, ideally with a considerable overlap of measured proteins. At the time of writing this thesis, I was not able to obtain access to any such dataset. However, it was recently announced that over 50,000 participants from the UKBB will have up to 1,536 proteins quantified using a recently developed protein measuring platform from Olink (Olink, 2020; UK Biobank, 2020). Several of the proteins highlighted in my analyses, such as kynureninase and r-spondin-3, will be measured, and this new data will thus offer great opportunities to confirm or reject my results in Chapter 3.

5.1.3.3. Functional follow-up of genes and proteins implicated in polycystic ovary syndrome

My work in Chapter 4 aimed to highlight potential disease-mediating genes for two obesity-related female reproductive diseases. Colocalisation analysis suggested potential effector genes at more than 30% of PCOS-associated risk loci, including several genes that have previously received little attention in PCOS research, such as *ZFP36L2*. A clear avenue for future research is to investigate these highlighted genes further, for example by using animal models and ovarian cell type lines, to determine and characterise their potential role in PCOS pathophysiology.

5.2. Conclusion

This thesis explores sex-specific causes, correlations, and consequences of obesity using genetic epidemiology. By combining large-scale sex-specific genetic, expression and proteomics data, I identify causal effects of proteins on obesity, and of obesity on disease risk, that differ between the

sexes. Furthermore, I highlight genes that may mediate the risk of a common obesity-related disease in women; PCOS. Thus, my work suggests several avenues that may lead to improved prediction, prevention, and treatment for obesity and obesity-related diseases in both men and women.

References

1.6 miljonerklubben (2020) *1,6 miljonerklubben: Hälsa och livskvalitet för alla kvinnor*. Available at: https://www.1.6miljonerklubben.com/ (Accessed: 23 October 2020).

Abete, I. *et al.* (2015) 'Association between different obesity measures and the risk of stroke in the EPIC Spanish cohort', *European Journal of Nutrition*, 54(3), pp. 365–375. doi: 10.1007/s00394-014-0716-x.

Abhishek, J. and Manuel, M. (2018) 'In aptamers they trust: the caveats of the SOMAscan biomarker discovery platform from SomaLogic', *Circulation*, 138(22), pp. 2482–2485. doi: 10.1161/CIRCULATIONAHA.118.036823.

Adams, C. D. and Neuhausen, S. L. (2019) 'Bi-directional Mendelian randomization of epithelial ovarian cancer and schizophrenia and uni-directional Mendelian randomization of schizophrenia on circulating 1- or 2-glycerophosphocholine metabolites', *Molecular Genetics and Metabolism Reports*, 21, p. 100539. doi: 10.1016/j.ymgmr.2019.100539.

Agudelo, L. Z. *et al.* (2014) 'Skeletal muscle PGC-1α1 modulates kynurenine metabolism and mediates resilience to stress-induced depression', *Cell*, 159(1), pp. 33–45. doi: 10.1016/j.cell.2014.07.051.

Akindele, M. O., Phillips, J. S. and Igumbor, E. U. (2016) 'The relationship between body fat percentage and body mass index in overweight and obese individuals in an urban African setting', *Journal of Public Health in Africa*, 7(1), p. 515. doi: 10.4081/jphia.2016.515.

Al-Ajmi, K. *et al.* (2018) 'Risk of breast cancer in the UK biobank female cohort and its relationship to anthropometric and reproductive factors', *PLoS One*, 13(7), p. e0201097. doi: 10.1371/journal.pone.0201097.

Alberts, B. *et al.* (2014) 'Protein function', in *Molecular Biology of the Cell*. 6th edn. Oxford: Taylor & Francis Group, pp. 109–169.

Aleman, A. *et al.* (2005) 'Bed rest during pregnancy for preventing miscarriage', *Cochrane Database of Systematic Reviews*, (2), p. CD003576. doi: 10.1002/14651858.cd003576.pub2.

American Diabetes Association (2010) 'Standards of medical care in diabetes-2010', *Diabetes Care*, 33(SUPPL. 1), pp. S11–S61. doi: 10.2337/dc10-S011.

Anand, B. K. and Brobeck, J. R. (1951) 'Hypothalamic control of food intake in rats and cats', *The Yale journal of biology and medicine*, 24(2), pp. 123–140.

Anderson, J. W. *et al.* (1999) 'Long-term weight maintenance after an intensive weight-loss program', *Journal of the American College of Nutrition*, 18(6), pp. 620–627. doi: 10.1080/07315724.1999.10718897.

Anderson, J. W. *et al.* (2001) 'Long-term weight-loss maintenance: a meta-analysis of US studies', *American Journal of Clinical Nutrition*, 74(5), pp. 579–84. doi: 10.1093/ajcn/74.5.579.

Di Angelantonio, E. *et al.* (2017) 'Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors', *Lancet*, 390(10110), pp. 2360–2371. doi: 10.1016/S0140-6736(17)31928-1.

Apple, F. S. *et al.* (2020) 'Sex-specific 99th percentile upper reference limits for high sensitivity cardiac troponin assays derived using a universal sample bank', *Clinical chemistry*, 66(3), pp. 434–444. doi: 10.1093/clinchem/hvz029.

Aschard, H. *et al.* (2015) 'Adjusting for heritable covariates can bias effect estimates in genome-wide association studies', *American Journal of Human Genetics*, 96(2), pp. 329–339. doi: 10.1016/j.ajhg.2014.12.021.

Ashwell, M., Cole, T. J. and Dixon, A. K. (1985) 'Obesity: new insight into the anthropometric classification of fat distribution shown by computed tomography', *BMJ*, 290(6483), pp. 1692–4. doi: 10.1136/bmj.290.6483.1692.

Astle, W. and Balding, D. J. (2009) 'Population structure and cryptic relatedness in genetic association studies', *Statistical Science*, 24(4), pp. 451–471. doi: 10.1214/09-STS307.

Astle, W. J. *et al.* (2016) 'The allelic landscape of human blood cell trait variation and links to common complex disease', *Cell*, 167(5), pp. 1415-1429.e19. doi: 10.1016/j.cell.2016.10.042.

Asunción, M. *et al.* (2000) 'A prospective study of the prevalence of the polycystic ovary syndrome in unselected caucasian women from Spain', *The Journal of Clinical Endocrinology & Metabolism*, 85(7), pp. 2434–2438. doi: 10.1210/jcem.85.7.6682.

Audrain-McGovern, J. and Benowitz, N. L. (2011) 'Cigarette smoking, nicotine, and body weight', *Clinical pharmacology and therapeutics*, 90(1), pp. 164–8. doi: 10.1038/clpt.2011.105.

Auguie, B. (2017) *gridExtra: Miscellaneous Functions for 'Grid' Graphics*. Available at:

https://cran.r-project.org/package=gridExtra (Accessed: 24 November 2020).

Auton, A. *et al.* (2015) 'A global reference for human genetic variation', *Nature*, 526(7571), pp. 68–74. doi: 10.1038/nature15393.

Aylward, A. *et al.* (2018) 'Shared genetic risk contributes to type 1 and type 2 diabetes etiology', *Human Molecular Genetics*, p. 285304. doi: 10.1093/hmg/ddy314.

Azziz, R. *et al.* (2016) 'Polycystic ovary syndrome', *Nature Reviews. Disease Primers*, 2, p. 16057. doi: 10.1038/nrdp.2016.57.

Babai, M. A. *et al.* (2016) 'Defining a BMI cut-off point for the iranian population: the shiraz heart study', *PLoS ONE*, 11(8). doi: 10.1371/journal.pone.0160639.

Baldini, G. and Phelan, K. D. (2019) 'The melanocortin pathway and control of appetite-progress and therapeutic implications', *Journal of Endocrinology*, 241(1), pp. R1–R33. doi: 10.1530/JOE-18-0596.

Balen, A. H. *et al.* (1995) 'Polycystic ovary syndrome: the spectrum of the disorder in 1741 patients', *Human Reproduction*, 10(8), pp. 2107–2111. doi: 10.1093/oxfordjournals.humrep.a136243.

Ball, C. B. *et al.* (2014) 'The RNA-binding protein, ZFP36L2, influences ovulation and oocyte maturation', *PLoS One*, 9(5), p. e97324. doi: 10.1371/journal.pone.0097324.

Band, G. and Marchini, J. (2020) *QCTOOL v2*. Available at: https://enkre.net/cgi-bin/code/qctool/dir?ci=trunk (Accessed: 4 January 2021).

Barroso, I. and McCarthy, M. I. (2019) 'The genetic basis of metabolic disease', *Cell*, 177(1), pp. 146–161. doi: 10.1016/j.cell.2019.02.024.

Barth, J. H. *et al.* (2010) 'Defining hyperandrogenism in polycystic ovary syndrome: Measurement of testosterone and androstenedione by liquid chromatography-tandem mass spectrometry and analysis by receiver operator characteristic plots', *European Journal of Endocrinology*, 162(3), pp. 611–615. doi: 10.1530/EJE-09-0741.

Bautista, L. E. *et al.* (2006) 'Estimation of bias in nongenetic observational studies using "Mendelian Triangulation"', *Annals of Epidemiology*, 16(9), pp. 675–680. doi: 10.1016/J.ANNEPIDEM.2006.02.001.

Beery, A. K. and Zucker, I. (2011) 'Sex bias in neuroscience and biomedical research', *Neuroscience and Biobehavioral Reviews*, 35(3), pp. 565–572. doi: 10.1016/j.neubiorev.2010.07.002.

Belanger, C. F. *et al.* (1978) 'The nurses' health study', *The American Journal of Nursing*, 78(6), pp. 1039–40.

Belmont, J. W. *et al.* (2005) 'A haplotype map of the human genome', *Nature*, 437(7063), pp. 1299–1320. doi: 10.1038/nature04226.

Bengtsson, H. (2017) *R.utils: Various Programming Utilities*. Available at: https://cran.r-project.org/package=R.utils (Accessed: 4 January 2021).

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), pp. 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x.

Berg, J. J. *et al.* (2019) 'Reduced signal for polygenic adaptation of height in UK Biobank', *Elife*, 8, p. e39725. doi: 10.7554/eLife.39725.

Berisa, T. and Pickrell, J. K. (2016) 'Approximately independent linkage disequilibrium blocks in human populations', *Bioinformatics*, 32(2), pp. 283–285. doi: 10.1093/bioinformatics/btv546.

Bhatia, S. and Kleinjan, D. A. (2014) 'Disruption of long-range gene regulation in human genetic disease: a kaleidoscope of general principles, diverse mechanisms and unique phenotypic consequences', *Human Genetics*, 133(7), pp. 815–845. doi: 10.1007/s00439-014-1424-6.

Biase, F. H. (2017) 'Oocyte developmental competence: Insights from cross-species differential gene expression and human oocyte-specific functional gene networks', *OMICS*, 21(3), pp. 156–168. doi: 10.1089/omi.2016.0177.

Bishara, A. J. and Hittner, J. B. (2017) 'Confidence intervals for correlations when data are not normal', *Behaviour Research Methods*, 49(1), pp. 294–309. doi: 10.3758/s13428-016-0702-8.

Blanchette, S. *et al.* (2006) 'Circulating progesterone and obesity in men', *Hormone and Metabolic Research*, 38(5), pp. 330–335. doi: 10.1055/s-2006-925392.

Bliss, C. (1967) *Statistics in biology: Statistical methods for research in the natural sciences*. New York: McGraw-Hill.

Bobadilla, J. L. *et al.* (2002) 'Cystic fibrosis: a worldwide analysis of CFTR mutations - correlation with incidence data and application to screening', *Human Mutation*, 19(6), pp. 575–606. doi: 10.1002/humu.10041.

Boef, A. G. C., Dekkers, O. M. and Le Cessie, S. (2015) 'Mendelian randomization studies: a review of the approaches used and the quality of reporting', *International Journal of Epidemiology*, 44(2), pp. 496–511. doi: 10.1093/ije/dyv071.

de Bold, A. J. *et al.* (1981) 'A rapid and potent natriuretic response to intravenous injection of atrial myocardial extract in rats', *Life Sciences*, 28(1), pp. 89–94. doi: 10.1016/0024-3205(81)90370-2.

Bonferroni, C. (1936) 'Teoria statistica delle classi e calcolo delle probabilita', *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, 8, pp. 3–62.

Boots, C. and Stephenson, M. D. (2011) 'Does obesity increase the risk of miscarriage in spontaneous conception: a systematic review', *Seminars in Reproductive Medicine*, 29(6), pp. 507–513. doi: 10.1055/s-0031-1293204.

Boots, L. R. *et al.* (1998) 'Measurement of total serum testosterone levels using commercially available kits: High degree of between-kit variability', *Fertility and Sterility*, 69(2), pp. 286–292. doi: 10.1016/S0015-0282(97)00464-0.

Bowden, J. *et al.* (2016) 'Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator', *Genetic epidemiology*, 40(4), pp. 304–14. doi: 10.1002/gepi.21965.

Bowden, J., Davey Smith, G. and Burgess, S. (2015) 'Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression', *International Journal of Epidemiology*, 44(2), pp. 512–525. doi: 10.1093/ije/dyv080.

Bozdag, G. *et al.* (2016) 'The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis', *Human Reproduction*, 31(12), pp. 2841–2855. doi: 10.1093/humrep/dew218.

Bradbury, K. E. *et al.* (2017) 'Association between physical activity and body fat percentage, with adjustment for BMI: a large cross-sectional analysis of UK Biobank', *BMJ Open*, 7(3), p. e011843. doi: 10.1136/bmjopen-2016-011843.

van den Brandt, P. A. *et al.* (2000) 'Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk', *American Journal of Epidemiology*, 152(6), pp. 514–527. doi: 10.1093/aje/152.6.514.

Braun, M. M., Stevens, W. A. and Barstow, C. (2018) 'Stable coronary artery disease: Treatment', *American Family Physician*, 97(6), pp. 376–384.

Bretherick, A. D. *et al.* (2020) 'Linking protein to phenotype with Mendelian Randomization detects 38 proteins with causal roles in human diseases and traits', *PLoS Genetics*, 16(7), p. e1008785. doi: 10.1371/journal.pgen.1008785.

Broad (2014) *SNPsnap*. Available at: https://data.broadinstitute.org/mpg/snpsnap/documentation.html (Accessed: 10 February 2021).

Brøchner-Mortensen, J., Rickers, H. and Balslev, I. (1980) 'Renal function and body composition before and after intestinal bypass operation in obese patients', *Scandinavian Journal of Clinical and Laboratory Investigation*, 40(8), pp. 695–702. doi: 10.3109/00365518009095584.

Broder-Fingert, S. *et al.* (2009) 'Evaluation of adolescents for polycystic ovary syndrome in an urban population', *Journal of Clinical Research in Pediatric Endocrinology*, 1(4), pp. 188–193. doi: 10.4274/jcrpe.v1i4.50.

Brower, M. A. *et al.* (2019) 'Bidirectional Mendelian randomization to explore the causal relationships between body mass index and polycystic ovary syndrome', *Human Reproduction*, 34(1), pp. 127–136. doi: 10.1093/humrep/dey343.

Brown, G. R. *et al.* (2015) 'Gene: a gene-centered information resource at NCBI', *Nucleic Acids Research*, 43(Database issue), pp. D36-42. doi: 10.1093/nar/gku1055.

Brumpton, B. *et al.* (2020) 'Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses', *Nature Communications*, 11(1), p. 3519. doi: 10.1038/s41467-020-17117-4.

Bulik-Sullivan, B. *et al.* (2015) 'An atlas of genetic correlations across human diseases and traits', *Nature genetics*, 47(11), pp. 1236–1241. doi: 10.1101/014498.

Bulik-Sullivan, B. and Finucane, H. (2017a) *FAQ*. Available at: https://github.com/bulik/ldsc/wiki/FAQ (Accessed: 16 February 2021).

Bulik-Sullivan, B. and Finucane, H. (2017b) *Heritability and Genetic Correlation*. Available at:

https://github.com/bulik/ldsc/wiki/Heritability-and-Genetic-Correlation (Accessed: 16 February 2021).

Bulik-Sullivan, B. K. *et al.* (2015) 'LD Score regression distinguishes confounding from polygenicity in genome-wide association studies', *Nature Genetics*, 47, pp. 291–295. doi: 10.1038/ng.3211.

Buniello, A. *et al.* (2019) 'The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019', *Nucleic Acids Research*, 47(D1), pp. D1005–D1012. doi: 10.1093/nar/gky1120.

Burgess, S. *et al.* (2015) 'Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors', *European Journal of Epidemiology*, 30(7), pp. 543–52. doi: 10.1007/s10654-015-0011-z.

Burgess, S. *et al.* (2017) 'Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants', *Epidemiology*, 28(1), pp. 30–42. doi: 10.1097/EDE.0000000000000559.

Burgess, S., Foley, C. N., *et al.* (2020) 'A robust and efficient method for Mendelian randomization with hundreds of genetic variants', *Nature Communications*, 11(1), p. 376. doi: 10.1038/s41467-019-14156-4.

Burgess, S., Davey Smith, G., *et al.* (2020) 'Guidelines for performing Mendelian randomization investigations', *Wellcome Open Research*, 4. doi: 10.12688/wellcomeopenres.15555.2.

Burgess, S., Butterworth, A. and Thompson, S. G. (2013) 'Mendelian randomization analysis with multiple genetic variants using summarized data', *Genetic Epidemiology*, 37(7), pp. 658–65. doi: 10.1002/gepi.21758.

Burgess, S., Davies, N. M. and Thompson, S. G. (2016) 'Bias due to participant overlap in two-sample Mendelian randomization', *Genetic Epidemiology*, 40(7), pp. 597–608. doi: 10.1002/gepi.21998.

Burgess, S., Dudbridge, F. and Thompson, S. G. (2016) 'Combining information on multiple instrumental variables in Mendelian randomization: Comparison of allele score and summarized data methods', *Statistics in Medicine*, 35(11), pp. 1880–906. doi: 10.1002/sim.6835.

Burgess, S., Small, D. S. and Thompson, S. G. (2017) 'A review of instrumental variable

estimators for Mendelian randomization', *Statistical Methods in Medical Research*, pp. 2333–2355. doi: 10.1177/0962280215597579.

Burgess, S. and Thompson, S. G. (2011) 'Avoiding bias from weak instruments in Mendelian randomization studies', *International Journal of Epidemiology*, 40(3), pp. 755–764. doi: 10.1093/ije/dyr036.

Burgess, S. and Thompson, S. G. (2013) 'Use of allele scores as instrumental variables for Mendelian randomization', *International Journal of Epidemiology*, 42(4), pp. 1134–1144. doi: 10.1093/ije/dyt093.

Burgess, S. and Thompson, S. G. (2015) *Mendelian Randomization - methods for using genetic variants in causal estimation*. 1st edn. Boca Raton, FL, USA: CRC Press, Taylor & Francis Group, Chapman and Hall, pp. 4, 6, 12, 20, 29, 31, 34–37, 39, 41, 45–52, 56, 58.

Butler, D. (2001) 'Epidemiology set to get fast-track treatment', *Nature*, 414(6860), p. 139. doi: 10.1038/35102713.

Buvinić, M. *et al.* (2011) 'Gender differentials in health', in Jamison, D. T. et al. (eds) *Disease Control Priorities in Developing Countries*. Washington (DC): World Bank.

Bycroft, C. *et al.* (2018) 'The UK Biobank resource with deep phenotyping and genomic data', *Nature*, 562(7726), pp. 203–209. doi: 10.1038/s41586-018-0579-z.

Çaliskan, M. *et al.* (2019) 'Genetic and epigenetic fine mapping of complex trait associated loci in the human liver', *American Journal of Human Genetics*, 105(1), pp. 89–107. doi: 10.1016/j.ajhg.2019.05.010.

Campbell, C. D. *et al.* (2005) 'Demonstrating stratification in a European American population', *Nature Genetics*, 37(8), pp. 868–872. doi: 10.1038/ng1607.

Campbell, J. M. *et al.* (2015) 'Paternal obesity negatively affects male fertility and assisted reproduction outcomes: a systematic review and meta-analysis', *Reproductive BioMedicine Online*, 31(5), pp. 593–604. doi: 10.1016/J.RBMO.2015.07.012.

Canadian Institutes of Health Research (2019) *How to integrate sex and gender into research*. Available at: https://cihr-irsc.gc.ca/e/50836.html (Accessed: 23 October 2020).

Cantwell, R. *et al.* (2011) 'Saving mothers' lives: Reviewing maternal deaths to make motherhood safer: 2006-2008', *BJOG: An International Journal of Obstetrics & Gynaecology*, 118, pp. 1–203.

doi: 10.1111/j.1471-0528.2010.02847.x.

CARDIoGRAMplusC4D Consortium (2020) *Data downloads*. Available at: http://www.cardiogramplusc4d.org/data-downloads/ (Accessed: 2 February 2021).

Carmo-Silva, S. and Cavadas, C. (2017) 'Hypothalamic dysfunction in obesity and metabolic disorders', in Letra, L. and Seiça, R. (eds) *Obesity and brain function*. Cham, p. 86. doi: 10.1007/978-3-319-63260-5.

Carreras-Torres, R. *et al.* (2017) 'Obesity, metabolic factors and risk of different histological types of lung cancer: a Mendelian randomization study', *PLoS One*, 12(6), p. e0177875. doi: 10.1371/journal.pone.0177875.

Carter, A. R. *et al.* (2019) 'Understanding the consequences of education inequality on cardiovascular disease: Mendelian randomisation study', *BMJ*, 365, p. l1855. doi: 10.1136/bmj.l1855.

Cartier, A. *et al.* (2009) 'Age-related differences in inflammatory markers in men: Contribution of visceral adiposity', *Metabolism: Clinical and Experimental*. Metabolism, 58(10), pp. 1452–1458. doi: 10.1016/j.metabol.2009.04.025.

Carvalho-Silva, D. *et al.* (2019) 'Open targets platform: New developments and updates two years on', *Nucleic Acids Research*. Narnia, 47(D1), pp. D1056–D1065. doi: 10.1093/nar/gky1133.

Cavalcante, M. B. *et al.* (2019) 'Obesity and recurrent miscarriage: a systematic review and meta-analysis', *Journal of Obstetrics and Gynaecology Research*, 45(1), pp. 30–38. doi: 10.1111/jog.13799.

Cavazos, T. B. and Witte, J. S. (2021) 'Inclusion of variants discovered from diverse populations improves polygenic risk score transferability', *Human Genetics and Genomics Advances*, 2(1), p. 100017. doi: 10.1016/j.xhgg.2020.100017.

Celis-Morales, C. A. *et al.* (2019) 'The combination of physical activity and sedentary behaviors modifies the genetic predisposition to obesity', *Obesity*, 27(4), pp. 653–661. doi: 10.1002/oby.22417.

Censin, J. C. *et al.* (2017) 'Childhood adiposity and risk of type 1 diabetes: a Mendelian randomization study', *PLOS Medicine*, 14(8), p. e1002362. doi: 10.1371/journal.pmed.1002362.

Censin, J. C., Peters, S. A. E., *et al.* (2019) 'Causal relationships between obesity and the leading

causes of death in women and men', *PLoS Genetics*, 15(10), p. e1008405. doi: 10.1371/journal.pgen.1008405.

Censin, J. C., Bovijn, J., *et al.* (2019) 'Causal relevance of obesity on the leading causes of death in women and men: a Mendelian randomization study', *bioRxiv*, p. 523217. doi: 10.1101/523217.

Censin, J. C. *et al.* (2020) 'Colocalization highlights genes in hypothalamic-pituitary-gonadal axis as potentially mediating polycystic ovary syndrome risk', *bioRxiv*, p. 2020.01.10.901116. doi: 10.1101/2020.01.10.901116.

Censin, J. C. *et al.* (2021) 'Colocalization analysis of polycystic ovary syndrome to identify potential disease-mediating genes and proteins', *European Journal of Human Genetics*. doi: 10.1038/s41431-021-00835-8.

Chakraborty, A. *et al.* (2015) 'Neil2-null mice accumulate oxidized DNA bases in the transcriptionally active sequences of the genome and are susceptible to innate inflammation', *Journal of Biological Chemistry*, 290(41), pp. 24636–24648. doi: 10.1074/jbc.M115.658146.

Chalasani, N. *et al.* (2012) 'The diagnosis and management of non-alcoholic fatty liver disease: Practice guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association', *Hepatology*, 55(6), pp. 2005–2023. doi: 10.1002/hep.25762.

Chan, E. W. *et al.* (2013) 'Efficacy and safety of lorcaserin in obese adults: a meta-analysis of 1-year randomized controlled trials (RCTs) and narrative review on short-term RCTs', *Obesity Reviews*, 14(5), pp. 383–392. doi: 10.1111/obr.12015.

Chang, C. (2021) *Association analysis*. Available at: https://www.cog-genomics.org/plink/2.0/assoc#glm_footnote1 (Accessed: 15 February 2021).

Chen, H., Fu, J. and Huang, W. (2016) 'Dopamine agonists for preventing future miscarriage in women with idiopathic hyperprolactinemia and recurrent miscarriage history', *Cochrane Database of Systematic Reviews*, 2016(7), p. CD008883. doi: 10.1002/14651858.CD008883.pub2.

Chen, Z.-J. *et al.* (2011) 'Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3', *Nature Genetics*, 43(1), pp. 55–59. doi: 10.1038/ng.732.

Chiolero, A. *et al.* (2008) 'Consequences of smoking for body weight, body fat distribution, and

insulin resistance', *The American Journal of Clinical Nutrition*, 87(4), pp. 801–809. doi: 10.1093/ajcn/87.4.801.

Cho, L. W. *et al.* (2006) 'The LH/FSH ratio has little use in diagnosing polycystic ovarian syndrome', *Annals of Clinical Biochemistry*, 43(Pt 3), pp. 217–219. doi: 10.1258/000456306776865188.

Chong, M. *et al.* (2019) 'Novel drug targets for ischemic stroke identified through Mendelian Randomization analysis of the blood proteome', *Circulation*, 140(10), pp. 819–830. doi: 10.1161/CIRCULATIONAHA.119.040180.

Chowdhury, I. *et al.* (2017) 'Gonadotropin-fependent neuregulin-1 signaling regulates female rat ovarian granulosa cell survival', *Endocrinology*, 158(10), pp. 3647–3660. doi: 10.1210/en.2017-00065.

Christensen, R. *et al.* (2007) 'Efficacy and safety of the weight-loss drug rimonabant: a meta-analysis of randomised trials', *Lancet*, 370(9600), pp. 1706–1713. doi: 10.1016/S0140-6736(07)61721-8.

Christiansen, O. B. *et al.* (1990) 'Idiopathic recurrent spontaneous abortion: Evidence of a familial predisposition', *Acta Obstetricia et Gynecologica Scandinavica*, 69(7–8), pp. 597–601. doi: 10.3109/00016349009028702.

Chun, S. *et al.* (2017) 'Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types', *Nature Genetics*, 49(4), pp. 600–605. doi: 10.1038/ng.3795.

Church, C. *et al.* (2010) 'Overexpression of Fto leads to increased food intake and results in obesity', *Nature Genetics*, 42(12), pp. 1086–1092. doi: 10.1038/ng.713.

Claussnitzer, M. *et al.* (2015) 'FTO obesity variant circuitry and adipocyte browning in humans', *New England Journal of Medicine*, 373(10), pp. 895–907. doi: 10.1056/nejmoa1502214.

Clayton, J. A. (2018) 'Applying the new SABV (sex as a biological variable) policy to research and clinical care', *Physiology and Behavior*, 187, pp. 2–5. doi: 10.1016/j.physbeh.2017.08.012.

Clearblue (2020) *All you need to know about hCG levels in early pregnancy*. Available at: https://uk.clearblue.com/pregnancy-tests/hcg (Accessed: 26 October 2020).

Cochran, W. (1954) 'The combination of estimates from different experiments', *Biometrics*, (10),

pp. 101–129.

Cohain, J. S., Buxbaum, R. E. and Mankuta, D. (2017) ‘Spontaneous first trimester miscarriage rates per woman among parous women with 1 or more pregnancies of 24 weeks or more’, *BMC Pregnancy and Childbirth*, 17(1), p. 437. doi: 10.1186/s12884-017-1620-1.

Collins, R. *et al.* (2002) ‘MRC/BHF heart protection study of antioxidant vitamin supplementation in 20 536 high-risk individuals: a randomised placebo-controlled trial’, *Lancet*, 360(9326), pp. 23–33. doi: 10.1016/S0140-6736(02)09328-5.

Condous, G. *et al.* (2006) ‘Failing pregnancies of unknown location: a prospective evaluation of the human chorionic gonadotrophin ratio’, *BJOG: An International Journal of Obstetrics and Gynaecology*, 113(5), pp. 521–527. doi: 10.1111/j.1471-0528.2006.00924.x.

Corbin, L. J. *et al.* (2016) ‘Body mass index as a modifiable risk factor for type 2 diabetes: Refining and understanding causal estimates using Mendelian randomisation’, *Diabetes*, 65(10), pp. 3002–3007. doi: 10.2337/db16-0418.

Cordina, M. *et al.* (2011) ‘Introduction of a single visit protocol in the management of selected patients with pregnancy of unknown location: a prospective study’, *BJOG: An International Journal of Obstetrics & Gynaecology*, 118(6), pp. 693–697. doi: 10.1111/j.1471-0528.2011.02893.x.

Crane, J. D. *et al.* (2015) ‘Inhibiting peripheral serotonin synthesis reduces obesity and metabolic dysfunction by promoting brown adipose tissue thermogenesis’, *Nature Medicine*, 21(2), pp. 166–172. doi: 10.1038/nm.3766.

Creemers, J. W. M. *et al.* (2012) ‘Heterozygous mutations causing partial prohormone convertase 1 deficiency contribute to human obesity’, *Diabetes*, 61(2), pp. 383–390. doi: 10.2337/db11-0305.

Cumming, G. P. *et al.* (2007) ‘The emotional burden of miscarriage for women and their partners: Trajectories of anxiety and depression over 13 months’, *BJOG: An International Journal of Obstetrics and Gynaecology*, 114(9), pp. 1138–1145. doi: 10.1111/j.1471-0528.2007.01452.x.

Curtis, D. (2018) ‘Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia’, *Psychiatric Genetics*, 28(5), pp. 85–89. doi: 10.1097/YPG.0000000000000206.

Dagenais, G. R. *et al.* (2005) ‘Prognostic impact of body weight and abdominal obesity in women

and men with cardiovascular disease', *American Heart Journal*, 149(1), pp. 54–60. doi: 10.1016/J.AHJ.2004.07.009.

Dale, C. E. *et al.* (2017) 'Causal associations of adiposity and body fat distribution with coronary heart disease, stroke subtypes, and type 2 diabetes mellitus: A Mendelian randomization analysis', *Circulation*, 135(24), pp. 2373–2388. doi: 10.1161/CIRCULATIONAHA.116.026560.

Damone, A. L. *et al.* (2019) 'Depression, anxiety and perceived stress in women with and without PCOS: a community-based study', *Psychological Medicine*, 49(9), pp. 1510–1520. doi: 10.1017/S0033291718002076.

Das, S. R. *et al.* (2005) 'Impact of body mass and body composition on circulating levels of natriuretic peptides: Results from the Dallas Heart Study', *Circulation*, 112(14), pp. 2163–2168. doi: 10.1161/CIRCULATIONAHA.105.555573.

Davey Smith, G. and Ebrahim, S. (2003) '"Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease?', *International Journal of Epidemiology*, 32(1), pp. 1–22. doi: 10.1093/ije/dyg070.

Davey Smith, G. and Hemani, G. (2014) 'Mendelian randomization: genetic anchors for causal inference in epidemiological studies.', *Human molecular genetics*, 23(R1), pp. R89-98. doi: 10.1093/hmg/ddu328.

Davey Smith, G., Paternoster, L. and Relton, C. (2017) 'When will Mendelian randomization become relevant for clinical practice and public health?', *JAMA*, 317(6), p. 589. doi: 10.1001/jama.2016.21189.

Davies, N. M., Holmes, M. V and Davey Smith, G. (2018) 'Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians', *BMJ*, 362, p. k601. doi: 10.1136/BMJ.K601.

Day, F. *et al.* (2018) 'Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria', *PLOS Genetics*, 14(12), p. e1007813. doi: 10.1371/journal.pgen.1007813.

Day, F. R. *et al.* (2015) 'Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome', *Nature Communications*, 6, p. 8464. doi: 10.1038/ncomms9464.

Day, F. R. *et al.* (2016) 'A robust example of collider bias in a genetic association study', *American Journal of Human Genetics*, 98(2), pp. 392–393. doi: 10.1016/j.ajhg.2015.12.019.

Demerath, E. W. *et al.* (2007) 'Anatomical patterning of visceral adipose tissue: Race, sex, and age variation', *Obesity*, 15(12), pp. 2984–2993. doi: 10.1038/oby.2007.356.

Deng, S.-N. *et al.* (2019) 'Long-term NMDAR antagonism correlates weight loss with less eating', *Frontiers in Psychiatry*, 10, p. 15. doi: 10.3389/fpsyt.2019.00015.

Denny, J. C. *et al.* (2013) 'Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data', *Nature Biotechnology*, 31(12), pp. 1102–1110. doi: 10.1038/nbt.2749.

Despres, J. P. *et al.* (1989) 'Role of deep abdominal fat in the association between regional adipose tissue distributioin and glucose tolerance in obese women', *Diabetes*, 38(3), pp. 304–309. doi: 10.2337/diab.38.3.304.

DeVon, H. A. *et al.* (2008) 'Symptoms across the continuum of acute coronary syndromes: Differences between women and men', *American Journal of Critical Care*, 17(1), pp. 14–24. doi: 10.4037/ajcc2008.17.1.14.

Dhana, K. *et al.* (2016) 'Body shape index in comparison with other anthropometric measures in prediction of total and cause-specific mortality', *Journal of epidemiology and community health*, 70(1), pp. 90–6. doi: 10.1136/jech-2014-205257.

DIAGRAM (2020) *Data download*. Available at: https://diagram-consortium.org/downloads.html (Accessed: 2 February 2021).

Dimou, N. L. *et al.* (2019) 'Sex hormone binding globulin and risk of breast cancer: a Mendelian randomization study', *Internation Journal of Epidemiology*, 48(3), pp. 807–816. doi: 10.1093/ije/dyz107.

Dina, C. *et al.* (2007) 'Variation in FTO contributes to childhood obesity and severe adult obesity', *Nature Genetics*, 39(6), pp. 724–726. doi: 10.1038/ng2048.

Doebler, P. (2017) *mada: Meta-analysis of diagnostic accuracy*. Available at: https://cran.r-project.org/package=mada%0A (Accessed: 24 November 2020).

Dombrowski, S. U. *et al.* (2014) 'Long term maintenance of weight loss with non-surgical interventions in obese adults: Systematic review and meta-analyses of randomised controlled

trials', *BMJ*, 348, p. g2646. doi: 10.1136/bmj.g2646.

Donovan, M. K. R. *et al.* (2020) 'Cellular deconvolution of GTEx tissues powers discovery of disease and cell-type associated regulatory variants', *Nature Communications*, 11(1), p. 955. doi: https://doi.org/10.1038/s41467-020-14561-0.

Dowle, M. and Srinivasan, A. (2020) *data.table: Extension of `data.frame`*. Available at: https://cran.r-project.org/web/packages/data.table/index.html (Accessed: 4 December 2020).

Dragulescu, A. A. and Arendt, C. (2018) *xlsx: Read, write, format Excel 2007 and Excel 97/2000/XP/2003 files*. Available at: https://cran.r-project.org/package=xlsx (Accessed: 5 March 2021).

Duncan, L. E. *et al.* (2018) 'Analysis of polygenic score usage and performance in diverse human populations', *bioRxiv*, p. 398396. doi: 10.1101/398396.

Durinck, S. *et al.* (2005) 'BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis', *Bioinformatics*, 21(16), pp. 3439–3440. doi: 10.1093/bioinformatics/bti525.

Durinck, S. *et al.* (2009) 'Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt', *Nature Protocols*, 4(8), pp. 1184–1191. doi: 10.1038/nprot.2009.97.

Eastwood, S. V *et al.* (2016) 'Algorithms for the capture and adjudication of prevalent and incident diabetes in UK Biobank', *PLoS One*, 11(9), p. e0162388. doi: 10.1371/journal.pone.0162388.

El-Kabbani, O., Dhagat, U. and Hara, A. (2011) 'Inhibitors of human 20α-hydroxysteroid dehydrogenase (AKR1C1)', *Journal of Steroid Biochemistry and Molecular Biology*, 125(1–2), pp. 105–111. doi: 10.1016/j.jsbmb.2010.10.006.

Emdin, C. A. *et al.* (2017) 'Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes, and coronary heart disease', *JAMA*, 317(6), pp. 626–634. doi: 10.1001/jama.2016.21042.

Emdin, C. A. *et al.* (2018) 'Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease', *Nature Communications*, 9, p. 1613. doi: 10.1038/s41467-018-03911-8.

ENCODE Project Consortium (2011) 'A user's guide to the encyclopedia of DNA elements

(ENCODE)', *PLoS Biology*, 9(4), p. e1001046. doi: 10.1371/journal.pbio.1001046.

Estes, C. *et al.* (2018) 'Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease', *Hepatology*, 67(1), pp. 123–133. doi: 10.1002/hep.29466.

European Medicines Agency (2013) *Refusal of the marketing authorisation for Qsiva (phentermine / topiramate)*. Available at: https://www.ema.europa.eu/en/documents/smop-initial/questions-answers-refusal-marketing-authorisation-qsiva-phentermine/topiramate_en.pdf.

Fabregat, A. *et al.* (2018) 'The Reactome pathway knowledgebase', *Nucleic Acids Research*, 46(D1), pp. D649–D655. doi: 10.1093/nar/gkx1132.

Farooqi, I. S. *et al.* (2000) 'Dominant and recessive inheritance of morbid obesity associate with melanocortin 4 receptor deficiency', *Journal of Clinical Investigation*, 106(2), pp. 271–279. doi: 10.1172/JCI9397.

Farooqi, I. S. *et al.* (2003) 'Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene', *New England Journal of Medicine*, 348(12), pp. 1085–1095. doi: 10.1056/NEJMoa022050.

Fauser, B. C. and Van Heusden, A. M. (1997) 'Manipulation of human ovarian function: Physiological concepts and clinical consequences', *Endocrine Reviews*, 18(1), pp. 71–106. doi: 10.1210/edrv.18.1.0290.

Fauser, B. C. J. M. *et al.* (2012) 'Consensus on women's health aspects of polycystic ovary syndrome (PCOS): the Amsterdam ESHRE/ASRM-Sponsored 3rd PCOS Consensus Workshop Group', *Fertility and Sterility*, 97(1), pp. 28-38.e25. doi: 10.1016/j.fertnstert.2011.09.024.

FDA-NIH Biomarker Working Group (2016) *BEST (Biomarkers , EndpointS, and other Tools)*. Bethesda: Food and Drug Administration (US). Available at: https://www.ncbi.nlm.nih.gov/books/NBK326791/ (Accessed: 13 October 2020).

Felix, J. F. *et al.* (2016) 'Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index', *Human Molecular Genetics*, 25(2), pp. 389–403. doi: 10.1093/hmg/ddv472.

Fischer, J. *et al.* (2009) 'Inactivation of the Fto gene protects from obesity', *Nature*, 458(7240), pp. 894–898. doi: 10.1038/nature07848.

Fisher, R. A. (1915) ‘Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population’, *Biometrika*, 10(4), pp. 507–521.

Fisher, R. A. (1921) ‘On the “probable error” of a coefficient of correlation seduced from a small sample’, *Metron*, (1), pp. 3–32.

Foley, C. N. *et al.* (2021) ‘A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits’, *Nature Communications*, 12, p. 764. doi: 10.1038/s41467-020-20885-8.

Folkersen, L. *et al.* (2020) ‘Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals’, *Nature Metabolism*, 2, pp. 1135–1148. doi: https://doi.org/10.1038/s42255-020-00287-2.

Food and Drug Administration (1977) *General considerations for the clinical evaluation of drugs*. Available at: https://www.fda.gov/media/71495/download.

Food and Drug Administration (1993) ‘Guidelines for the study and evaluation of gender differences in the clinical evaluation of drugs; notice’, *Federal Register*, 58(139), pp. 39406–39416.

Fox, C. S. *et al.* (2007) ‘Abdominal visceral and subcutaneous adipose tissue compartments: Association with metabolic risk factors in the framingham heart study’, *Circulation*, 116(1), pp. 39–48. doi: 10.1161/CIRCULATIONAHA.106.675355.

Franceschini, N. *et al.* (2018) ‘GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes’, *Nature Communications*, 9(1), p. 5141. doi: 10.1038/s41467-018-07340-5.

Frank, L. K. *et al.* (2013) ‘Measures of general and central obesity and risk of type 2 diabetes in a Ghanaian population’, *Tropical Medicine and International Health*, 18(2), pp. 141–151. doi: 10.1111/tmi.12024.

Frayling, T. M. *et al.* (2007) ‘A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity’, *Science*, 316(5826), pp. 889–894. doi: 10.1126/science.1141634.

Free Software Foundation (2020) *GNU operating system*. Available at: https://www.gnu.org/ (Accessed: 24 November 2020).

Fried, S. K., Bunkin, D. A. and Greenberg, A. S. (1998) 'Omental and subcutaneous adipose tissues of obese subjects release interleukin-6: Depot difference and regulation by glucocorticoid', *The Journal of Clinical Endocrinology & Metabolism*, 83(3), pp. 847–850. doi: 10.1210/jcem.83.3.4660.

Fry, A. *et al.* (2017) 'Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population', *American Journal of Epidemiology*, 186(9), pp. 1026–1034. doi: 10.1093/aje/kwx246.

Fujioka, S. *et al.* (1987) 'Contribution of intra-abdominal fat accumulation to the impairment of glucose and lipid metabolism in human obesity', *Metabolism*, 36(1), pp. 54–59. doi: 10.1016/0026-0495(87)90063-1.

Fujitani, Y. *et al.* (2010) 'Pronounced adipogenesis and increased insulin sensitivity caused by overproduction of prostaglandin D2 in vivo', *The FEBS journal*, 277(6), pp. 1410–1419. doi: 10.1111/j.1742-4658.2010.07565.x.

Fulkerson, J. A. and French, S. A. (2003) 'Cigarette smoking for weight loss or control among adolescents: Gender and racial/ethnic differences', *The Journal of Adolescent Health*, 32(4), pp. 306–13. doi: 10.1016/s1054-139x(02)00566-9.

Gál, E. M. and Sherman, A. D. (1980) 'L-kynurenine: its synthesis and possible regulatory function in brain', *Neurochemical Research*, 5(3), pp. 223–239. doi: 10.1007/BF00964611.

Galaviz, K. I. *et al.* (2018) 'Global diabetes prevention interventions: a systematic review and network meta-analysis of the real-world impact on incidence, weight, and glucose', *Diabetes Care*, 41(7), pp. 1526–1534. doi: 10.2337/dc17-2222.

Galinsky, K. J. *et al.* (2016) 'Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia', *American Journal of Human Genetics*, 98(3), pp. 456–472. doi: 10.1016/j.ajhg.2015.12.022.

Gay, N. R. *et al.* (2020) 'Impact of admixture and ancestry on eQTL analysis and GWAS colocalization in GTEx', *Genome Biology*, 21(1), p. 233. doi: 10.1186/s13059-020-02113-0.

GBD 2015 Obesity Collaborators *et al.* (2017) 'Health effects of overweight and obesity in 195 countries over 25 Years', *The New England Journal of Medicine*, 377(1), pp. 13–27. doi: 10.1056/NEJMoa1614362.

Ge, T. *et al.* (2017) 'Phenome-wide heritability analysis of the UK Biobank', *PLoS Genetics*, 13(4), p. e1006711. doi: 10.1371/journal.pgen.1006711.

GEFOS Consortium (2018) *Data release 2018*. Available at: http://www.gefos.org/?q=content/data-release-2018 (Accessed: 3 February 2021).

Geller, S. E. *et al.* (2011) 'Inclusion, analysis, and reporting of sex and race/ethnicity in clinical trials: Have we made progress?', *Journal of Women's Health*, 20(3), pp. 315–320. doi: 10.1089/jwh.2010.2469.

Geller, S. E. *et al.* (2018) 'The more things change, the more they stay the same: a study to evaluate compliance with inclusion and assessment of women and minorities in randomized controlled trials', *Academic Medicine*, 93(4), pp. 630–635. doi: 10.1097/ACM.0000000000002027.

Geng, T. *et al.* (2018) 'Childhood BMI and adult type 2 diabetes, coronary artery diseases, chronic kidney disease, and cardiometabolic traits: A Mendelian Randomization analysis', *Diabetes care*, 41(5), pp. 1089–1096. doi: 10.2337/dc17-2141.

Gerken, T. *et al.* (2007) 'The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase', *Science*, 318(5855), pp. 1469–1472. doi: 10.1126/science.1151710.

Ghamari-Langroudi, M., Srisai, D. and Cone, R. D. (2011) 'Multinodal regulation of the arcuate/paraventricular nucleus circuit by leptin', *Proceedings of the National Academy of Sciences of the United States of America*, 108(1), pp. 355–360. doi: 10.1073/pnas.1016785108.

Giambartolomei, C. *et al.* (2014) 'Bayesian test for colocalisation between pairs of genetic association studies using summary statistics', *PLoS Genetics*, 10(5), p. e1004383. doi: 10.1371/journal.pgen.1004383.

Gibson-Helm, M. *et al.* (2017) 'Delayed diagnosis and a lack of information associated with dissatisfaction in women with polycystic ovary syndrome', *The Journal of Clinical Endocrinology & Metabolism*, 102(2), pp. 604–612. doi: 10.1210/jc.2016-2963.

Gloy, V. L. *et al.* (2013) 'Bariatric surgery versus non-surgical treatment for obesity: a systematic review and meta-analysis of randomised controlled trials', *BMJ*, 347, p. f5934. doi: 10.1136/bmj.f5934.

Glueck, C. J. *et al.* (2005) 'Obesity and extreme obesity, manifest by ages 20-24 years, continuing through 32-41 years in women, should alert physicians to the diagnostic likelihood of polycystic

ovary syndrome as a reversible underlying endocrinopathy', *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 122(2), pp. 206–212. doi: 10.1016/j.ejogrb.2005.03.010.

Gokhale, A. S. *et al.* (2019) 'Selective deletion of Eos (Ikzf4) in T-regulatory cells leads to loss of suppressive function and development of systemic autoimmunity', *Journal of Autoimmunity*, 105, p. 102300. doi: 10.1016/j.jaut.2019.06.011.

Gold, L. *et al.* (2010) 'Aptamer-based multiplexed proteomic technology for biomarker discovery', *PLoS One*, 5(12), p. e15004. doi: 10.1371/journal.pone.0015004.

Greenland, S. (2000) 'An introduction to instrumental variables for epidemiologists', *International journal of epidemiology*, 29(4), pp. 722–9. doi: 10.1093/ije/29.4.722.

Greenwood, E. A. *et al.* (2019) 'Depression over the lifespan in a population-based cohort of women with polycystic ovary syndrome: Longitudinal analysis', *Journal of Clinical Endocrinology and Metabolism*, 104(7), pp. 2809–2819. doi: 10.1210/jc.2019-00234.

Grinde, K. E. *et al.* (2019) 'Generalizing polygenic risk scores from europeans to hispanics/latinos', *Genetic Epidemiology*, 43(1), pp. 50–62. doi: 10.1002/gepi.22166.

GTEx Consortium (2017) 'Genetic effects on gene expression across human tissues', *Nature*, 550(7675), pp. 204–213. doi: 10.1038/nature24277.

GTEx Consortium (2020) *GTEx Portal*. Available at: https://gtexportal.org/home/ (Accessed: 4 December 2020).

Gudmundsdottir, V. *et al.* (2020) 'Circulating protein signatures and causal candidates for type 2 siabetes', *Diabetes*, 69(8), pp. 1843–1853. doi: 10.2337/db19-1070.

Guo, C. *et al.* (2019) 'Identification of putative effector genes across the GWAS Catalog using molecular quantitative trait loci from 68 tissues and cell types', *bioRxiv*, p. 808444. doi: 10.1101/808444.

Guo, H. *et al.* (2015) 'Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases', *Human Molecular Genetics*, 24(12), pp. 3305–3313. doi: 10.1093/hmg/ddv077.

Guo, Y. *et al.* (2016) 'Genetically predicted body mass index and breast cancer risk: Mendelian randomization analyses of data from 145,000 women of European descent', *PLoS Medicine*, 13(8), p. e1002105. doi: 10.1371/journal.pmed.1002105.

Halford, J. C. and Blundell, J. E. (2000) 'Separate systems for serotonin and leptin in appetite control', *Annals of Medicine*, 32(3), pp. 222–232. doi: 10.3109/07853890008998829.

Hall, K. D. *et al.* (2009) 'The progressive increase of food waste in America and its environmental impact', *PLoS ONE*, 4(11), p. e7940. doi: 10.1371/journal.pone.0007940.

Hall, M. E. *et al.* (2014) 'Obesity, hypertension, and chronic kidney disease', *International Journal of Nephrology and Renovascular Disease*, 7, pp. 75–88. doi: 10.2147/IJNRD.S39739.

Hanson, C. *et al.* (2014) 'Influence of diet and obesity on COPD development and outcomes', *International journal of chronic obstructive pulmonary disease*, 9, pp. 723–33. doi: 10.2147/COPD.S50111.

Harris, H. R. *et al.* (2019) 'Association between genetically predicted polycystic ovary syndrome and ovarian cancer: a Mendelian randomization study', *Internation Journal of Epidemiology*, 48(3), pp. 822–830. doi: https://doi.org/10.1093/ije/dyz113.

Hartwig, F. P., Smith, G. D. and Bowden, J. (2017) 'Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption', *International Journal of Epidemiology*, 46(6), pp. 1985–1998. doi: 10.1093/ije/dyx102.

Hauptman, J. B., Jeunet, F. S. and Hartmann, D. (1992) 'Initial studies in humans with the novel gastrointestinal lipase inhibitor Ro 18–0647 (tetrahydrolipstatin)', *The American Journal of Clinical Nutrition*, 55(1), pp. 309S-313S. doi: 10.1093/ajcn/55.1.309s.

Haworth, S. *et al.* (2019) 'Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis', *Nature Communications*, 10(1), p. 333. doi: 10.1038/s41467-018-08219-1.

Hayes, M. G. *et al.* (2015) 'Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations', *Nature Communications*, 6, p. 7502. doi: 10.1038/ncomms8502.

He, Z. *et al.* (2018) 'Plasma steroids, body composition, and fat distribution: Effects of age, sex, and exercise training', *International Journal of Obesity*, 42(7), pp. 1366–1377. doi: 10.1038/s41366-018-0033-1.

Heidari, S. *et al.* (2016) 'Sex and gender equity in research: Rationale for the SAGER guidelines and recommended use', *Research Integrity and Peer Review*, 1(1), p. 2. doi: 10.1186/s41073-016-

0007-6.

Hernandez-Carretero, A. *et al.* (2018) 'Obesity-induced changes in lipid mediators persist after weight loss', *International Journal of Obesity*, 42(4), pp. 728–736. doi: 10.1038/ijo.2017.266.

Heymsfield, S. B. and Wadden, T. A. (2017) 'Mechanisms, pathophysiology, and management of obesity', *New England Journal of Medicine*, 376(3), pp. 254–266. doi: 10.1056/nejmra1514009.

Hooper, L., Ness, A. R. and Smith, G. D. (2001) 'Antioxidant strategy for cardiovascular diseases', *Lancet*, pp. 1705–1706. doi: 10.1016/s0140-6736(00)04876-5.

Hormozdiari, F. *et al.* (2016) 'Colocalization of GWAS and eQTL Signals Detects Target Genes', *The American Journal of Human Genetics*, 99(6), pp. 1245–1260. doi: 10.1016/j.ajhg.2016.10.003.

Howie, B. N., Donnelly, P. and Marchini, J. (2009) 'A flexible and accurate genotype imputation method for the next generation of genome-wide association studies', *PLoS Genetics*, 5(6), p. e1000529. doi: 10.1371/journal.pgen.1000529.

Hu, G. *et al.* (2007) 'Body mass index, waist circumference, and waist-hip ratio on the risk of total and type-specific stroke', *Archives of Internal Medicine*, 167(13), p. 1420. doi: 10.1001/archinte.167.13.1420.

Huang, J. *et al.* (2015) 'Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel', *Nature Communications*, 6, pp. 1–9. doi: 10.1038/ncomms9111.

Huang, T. *et al.* (2015) 'Genetic predisposition to central obesity and risk of type 2 diabetes: Two independent cohort studies', *Diabetes care*, 38(7), pp. 1306–11. doi: 10.2337/dc14-3084.

Hwang, Y. C. *et al.* (2015) 'Visceral abdominal fat accumulation predicts the conversion of metabolically healthy obese subjects to an unhealthy phenotype', *International Journal of Obesity*, 39(9), pp. 1365–1370. doi: 10.1038/ijo.2015.75.

Ibrahim, M. M. (2010) 'Subcutaneous and visceral adipose tissue: Structural and functional differences', *Obesity Reviews*, pp. 11–18. doi: 10.1111/j.1467-789X.2009.00623.x.

Inagaki, A., Roset, R. and Petrini, J. H. J. (2016) 'Functions of the MRE11 complex in the development and maintenance of oocytes', *Chromosoma*, 125(1), pp. 151–162. doi: 10.1007/s00412-015-0535-8.

Ingelfinger, J. R. and Rosen, C. J. (2021) 'STEP 1 for effective weight control — another first

step?', *New England Journal of Medicine*. doi: 10.1056/NEJMe2101705.

International Consortium for Blood Pressure Genome-Wide Association Studies *et al.* (2011) 'Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk', *Nature*, 478(7367), pp. 103–9. doi: 10.1038/nature10405.

International Diabetes Federation (2006) *The IDF consensus worldwide definition of the metabolic syndrome*. Available at: https://www.idf.org/e-library/consensus-statements/60-idfconsensus-worldwide-definitionof-the-metabolic-syndrome.html.

International Schizophrenia Consortium, S. M. *et al.* (2009) 'Common polygenic variation contributes to risk of schizophrenia and bipolar disorder', *Nature*, 460(7256), pp. 748–52. doi: 10.1038/nature08185.

Jackson, R. S. *et al.* (1997) 'Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene', *Nature Genetics*, 16(3), pp. 303–306. doi: 10.1038/ng0797-303.

James, W. P. T. *et al.* (2010) 'Effect of sibutramine on cardiovascular outcomes in overweight and obese subjects', *New England Journal of Medicine*, 363(10), pp. 905–917. doi: 10.1056/NEJMoa1003114.

Jaroudi, S. *et al.* (2009) 'Expression profiling of DNA repair genes in human oocytes and blastocysts using microarrays', *Human Reproduction*, 24(10), pp. 2649–2655. doi: 10.1093/humrep/dep224.

Jarvis, D. *et al.* (2016) 'Mendelian randomisation analysis strongly implicates adiposity with risk of developing colorectal cancer', *British Journal of Cancer*, 115(2), pp. 266–272. doi: 10.1038/bjc.2016.188.

Joham, A. E. *et al.* (2015) 'Prevalence of infertility and use of fertility treatment in women with polycystic ovary syndrome: Data from a large community-based cohort study', *Journal of Women's Health*, 24(4), pp. 299–307. doi: 10.1089/jwh.2014.5000.

Jones, G. L. *et al.* (2008) 'Health-related quality of life measurement in women with polycystic ovary syndrome: a systematic review', *Human Reproduction Update*, 14(1), pp. 15–25. doi: 10.1093/humupd/dmm030.

Jurkovic, D., Overton, C. and Bender-Atik, R. (2013) 'Diagnosis and management of first trimester

miscarriage', *BMJ*, 346(7913), p. f3676. doi: 10.1136/bmj.f3676.

Kanaoka, Y. and Urade, Y. (2003) 'Hematopoietic prostaglandin D synthase', *Prostaglandins, Leukotriens and Essential Fatty Acids*, 69(2–3), pp. 163–167. doi: 10.1016/s0952-3278(03)00077-2.

Kanehisa, M. *et al.* (2019) 'New approach for understanding genome variations in KEGG', *Nucleic Acids Research*, 47(D1), pp. D590–D595. doi: 10.1093/nar/gky962.

Karakas, M. *et al.* (2018) 'Testosterone levels and type 2 diabetes—no correlation with age, differential predictive value in men and women', *Biomolecules*, 8(3), p. 76. doi: 10.3390/biom8030076.

Kassambara, A. (2020) *ggpubr: 'ggplot2' based publication ready plots*. Available at: https://github.com/kassambara/ggpubr/ (Accessed: 24 November 2020).

Katan, M. B. (1986) 'Apolipoprotein E isoforms, serum cholesterol, and cancer', *The Lancet*, pp. 507–508. doi: 10.1016/S0140-6736(86)92972-7.

Khan, E. *et al.* (2018) 'Differences in management and outcomes for men and women with ST-elevation myocardial infarction', *Medical Journal of Australia*, 209(3), pp. 118–123. doi: 10.5694/mja17.01109.

Khaw, K. T. *et al.* (2001) 'Relation between plasma ascorbic acid and mortality in men and women in EPIC-Norfolk prospective study: a prospective population study', *Lancet*, 357(9257), pp. 657–663. doi: 10.1016/S0140-6736(00)04128-3.

Khera, R. *et al.* (2016) 'Association of pharmacological treatments for obesity with weight loss and adverse events a systematic review and meta-analysis', *JAMA*, 315(22), pp. 2424–2434. doi: 10.1001/jama.2016.7602.

Khramtsova, E. A., Davis, L. K. and Stranger, B. E. (2019) 'The role of sex in the genomics of human complex traits', *Nature Reviews Genetics*, 20(3), pp. 173–190. doi: 10.1038/s41576-018-0083-1.

Kikas, T. *et al.* (2019) 'The effect of genetic variation on the placental transcriptome in humans', *Frontiers in Genetics*, 10, p. 550. doi: 10.3389/fgene.2019.00550.

Kim, J. *et al.* (2005) 'GDF11 controls the timing of progenitor cell competence in developing retina', *Science*, 308(5730), pp. 1927–1930. doi: 10.1126/science.1110175.

Kimenai, D. M. *et al.* (2018) 'Sex-specific versus overall clinical decision limits for cardiac troponin i and T for the diagnosis of acute myocardial infarction: a systematic review', *Clinical Chemistry*, 64(7), pp. 1034–1043. doi: 10.1373/clinchem.2018.286781.

Knochenhauer, E. S. *et al.* (1998) 'Prevalence of the polycystic ovary syndrome in unselected black and white women of the southeastern United States: a prospective study', *The Journal of Clinical Endocrinology & Metabolism*, 83(9), pp. 3078–3082. doi: 10.1210/jcem.83.9.5090.

Knowler, W. C. *et al.* (1988) 'Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture', *American Journal of Human Genetics*, 43(4), pp. 520–526.

Kolte, A. M. *et al.* (2011) 'A genome-wide scan in affected sibling pairs with idiopathic recurrent miscarriage suggests genetic linkage', *Molecular Human Reproduction*, 17(6), pp. 379–385. doi: 10.1093/molehr/gar003.

Kong, A. *et al.* (2018) 'The nature of nurture: Effects of parental genotypes', *Science*, 359(6374), pp. 424–428. doi: 10.1126/science.aan6877.

Kovesdy, C. P. *et al.* (2017) 'Obesity and kidney disease: Hidden consequences of the epidemic', *Canadian journal of kidney health and disease*, 4, p. 2054358117698669. doi: 10.1177/2054358117698669.

Krijger, P. H. L. and De Laat, W. (2016) 'Regulation of disease-associated gene expression in the 3D genome', *Nature Reviews Molecular Cell Biology*, 17(12), pp. 771–782. doi: 10.1038/nrm.2016.138.

Krishnan, A. and Muthusami, S. (2017) 'Hormonal alterations in PCOS and its influence on bone metabolism', *Journal of Endocrinology*, 232(2), pp. R99–R113. doi: 10.1530/JOE-16-0405.

Kristjansson, R. P. *et al.* (2017) 'Sequence variant at 4q25 near PITX2 associates with appendicitis', *Scientific Reports*, 7(1), p. 3119. doi: 10.1038/s41598-017-03353-0.

Lagou, V. *et al.* (2021) 'Sex-dimorphic genetic effects and novel loci for fasting glucose and insulin variability', *Nature Communications*, 12(1), p. 24. doi: https://doi.org/10.1038/s41467-020-19366-9.

Laisk, T. *et al.* (2020) 'The genetic architecture of sporadic and multiple consecutive miscarriage', *Nature Communications*, 11(1), p. 5980. doi: 10.1038/s41467-020-19742-5.

Lander, E. S. *et al.* (2001) 'Initial sequencing and analysis of the human genome', *Nature*, 409(6822), pp. 860–921. doi: 10.1038/35057062.

Larsen, E. C. *et al.* (2013) 'New insights into mechanisms behind miscarriage', *BMC Medicine*, p. 154. doi: 10.1186/1741-7015-11-154.

Lawlor, D. A. *et al.* (2008) 'Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology', *Statistics in Medicine*, 27(8), pp. 1133–1163. doi: 10.1002/sim.3034.

Le, M. T. *et al.* (2019) 'Exploration of the role of anti-Mullerian hormone and LH/FSH ratio in diagnosis of polycystic ovary syndrome', *Clinical Endocrinology*, 90(4), pp. 579–585. doi: 10.1111/cen.13934.

Lee, H.-S. *et al.* (2018) 'Association of DKN2A/CDKN2B with inflammatory bowel disease in Koreans', *Journal of Gastroenterology and Hepatology*, 33(4), pp. 887–893. doi: 10.1111/jgh.14031.

Lee, S. K. (2018) 'Sex as an important biological variable in biomedical research', *BMB Reports*, pp. 167–173. doi: 10.5483/BMBRep.2018.51.4.034.

Li, C. *et al.* (2006) 'Sex differences in the relationships between BMI, WHR and incidence of cardiovascular disease: a population-based cohort study', *International Journal of Obesity*, 30(12), pp. 1775–1781. doi: 10.1038/sj.ijo.0803339.

Li, J. Y. *et al.* (2014) 'LGR4 and its ligands, R-spondin 1 and R-spondin 3, regulate food intake in the hypothalamus of male rats', *Endocrinology*, 155(2), pp. 429–440. doi: 10.1210/en.2013-1550.

Li, Yan *et al.* (2019) 'Multi-system reproductive metabolic disorder: Significance for the pathogenesis and therapy of polycystic ovary syndrome (PCOS)', *Life Sciences*, 228, pp. 167–175. doi: 10.1016/j.lfs.2019.04.046.

Lim, S. S. *et al.* (2013) 'The effect of obesity on polycystic ovary syndrome: a systematic review and meta-analysis', *Obesity Reviews*, 14(2), pp. 95–109. doi: 10.1111/j.1467-789X.2012.01053.x.

Lim, S. S. *et al.* (2019) 'Lifestyle changes in women with polycystic ovary syndrome', *Cochrane Database of Systematic Reviews*, 3(3), p. CD007506. doi: 10.1002/14651858.CD007506.pub4.

Liu, B., Pjanic, M., *et al.* (2018) 'Genetic regulatory mechanisms of smooth muscle cells map to coronary artery disease risk loci', *The American Journal of Human Genetics*, 103(3), pp. 377–388.

doi: 10.1016/J.AJHG.2018.08.001.

Liu, B., Calton, M. A., *et al.* (2018) 'Ocular disease mechanisms elucidated by genetics of human fetal retinal pigment epithelium gene expression', *bioRxiv*, p. 446799. doi: 10.1101/446799.

Liu, B., Gloudemans, M. and Montgomery, S. (2018) *LocusCompare: a tool to visualize pairs of association*. Available at: http://www.locuscompare.com/ (Accessed: 25 November 2020).

Liu, J.-P. (2006) 'The function of growth/differentiation factor 11 (Gdf11) in rostrocaudal patterning of the developing spinal cord', *Development*, 133(15), pp. 2865–2874. doi: 10.1242/dev.02478.

Liu, J. *et al.* (2019) 'The effect of liver enzymes on adiposity: a Mendelian randomization study', *Scientific Reports*, 9(1), pp. 1–10. doi: 10.1038/s41598-019-52489-8.

Liu, K. A. and Mager, N. A. D. (2016) 'Women's involvement in clinical trials: Historical perspective and future implications', *Pharmacy Practice*, 14(1), p. 708. doi: 10.18549/PharmPract.2016.01.708.

Liu, X.-M. *et al.* (2018) 'Loss of oocyte Rps26 in mice arrests oocyte growth and causes premature ovarian failure', *Cell Death & Disease*, 9(12), p. 1144. doi: 10.1038/s41419-018-1196-3.

Liu, Y. *et al.* (2017) 'The BMP4-Smad signaling pathway regulates hyperandrogenism development in a female mouse model', *Journal of Biological Chemistry*, 292(28), pp. 11740–11750. doi: 10.1074/jbc.M117.781369.

Locke, A. E. *et al.* (2015) 'Genetic studies of body mass index yield new insights for obesity biology', *Nature*, 518(7538), pp. 197–206. doi: 10.1038/nature14177.

Loh, N. Y. *et al.* (2020) 'RSPO3 impacts body fat distribution and regulates adipose cell biology in vitro', *Nature Communications*, 11(1), pp. 1–15. doi: 10.1038/s41467-020-16592-z.

Loh, P. R. *et al.* (2015) 'Efficient Bayesian mixed-model analysis increases association power in large cohorts', *Nature Genetics*, 47(3), pp. 284–290. doi: 10.1038/ng.3190.

Lopes-Ramos, C. M. *et al.* (2020) 'Sex differences in gene expression and regulatory networks across 29 human tissues', *Cell Reports*, 31(12), p. 107795. doi: 10.1016/j.celrep.2020.107795.

Lord, J. *et al.* (2006) 'The central issue? Visceral fat mass is a good marker of insulin resistance and metabolic disturbance in women with polycystic ovary syndrome', *BJOG: An International*

Journal of Obstetrics & Gynaecology, 113(10), pp. 1203–1209. doi: 10.1111/j.1471-0528.2006.00973.x.

Lotta, L. A. *et al.* (2018) 'Association of genetic variants related to gluteofemoral vs abdominal fat distribution with type 2 diabetes, coronary disease, and cardiovascular risk factors', *JAMA*, 320(24), p. 2553. doi: 10.1001/jama.2018.19329.

Lower, A. M. and Yovich, J. L. (1992) 'The value of serum levels of oestradiol, progesterone and β-human chorionic gonadotrophin in the prediction of early pregnancy loss', *Human Reproduction*, 7(5), pp. 711–717. doi: 10.1093/oxfordjournals.humrep.a137723.

Luo, H. *et al.* (2019) 'Growth differentiation factor 11 inhibits adipogenic differentiation by activating TGF-beta/Smad signalling pathway', *Cell Proliferation*, 52(4), p. e12631. doi: 10.1111/cpr.12631.

Lyall, D. M. *et al.* (2017) 'Association of body mass index with cardiometabolic disease in the UK Biobank', *JAMA Cardiology*, 2(8), p. 882. doi: 10.1001/jamacardio.2016.5804.

Macaluso, M. *et al.* (2010) 'A public health focus on infertility prevention, detection, and management', *Fertility and Sterility*, 93(1), pp. 16.e1-16.e10. doi: 10.1016/J.FERTNSTERT.2008.09.046.

Machiela, M. J. and Chanock, S. J. (2015) 'LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants', *Bioinformatics*, 31(21), pp. 3555–3557. doi: 10.1093/bioinformatics/btv402.

Madamanchi, C. *et al.* (2014) 'Obesity and natriuretic peptides, BNP and NT-proBNP: Mechanisms and diagnostic implications for heart failure', *International Journal of Cardiology*, 176(3), pp. 611–617. doi: 10.1016/j.ijcard.2014.08.007.

Mägi, R. *et al.* (2017) 'Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution', *Human Molecular Genetics*, 26(18), pp. 3639–3650. doi: 10.1093/hmg/ddx280.

Mahajan, A. *et al.* (2014) 'Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility', *Nature Genetics*, 46(3), pp. 234–244. doi: 10.1038/ng.2897.

Mahajan, A., Taliun, D., *et al.* (2018) 'Fine-mapping type 2 diabetes loci to single-variant

resolution using high-density imputation and islet-specific epigenome maps', *Nature Genetics*, 50(11), pp. 1505–1513. doi: 10.1038/s41588-018-0241-6.

Mahajan, A., Wessel, J., *et al.* (2018) 'Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes article', *Nature Genetics*, 50(4), pp. 559–571. doi: 10.1038/s41588-018-0084-1.

Mahto, A. (2019) *splitstackshape: Stack and reshape datasets after splitting concatenated values*. Available at: https://cran.r-project.org/package=splitstackshape (Accessed: 4 January 2021).

Mailand, N. *et al.* (2007) 'RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins', *Cell*, 131(5), pp. 887–900. doi: 10.1016/j.cell.2007.09.040.

Maino, A. *et al.* (2016) 'Pregnancy loss and risk of ischaemic stroke and myocardial infarction', *British Journal of Haematology*, 174(2), pp. 302–309. doi: 10.1111/bjh.14043.

Manichaikul, A. *et al.* (2010) 'Robust relationship inference in genome-wide association studies.', *Bioinformatics*, 26(22), pp. 2867–73. doi: 10.1093/bioinformatics/btq559.

Marchini, J. *et al.* (2007) 'A new multipoint method for genome-wide association studies by imputation of genotypes', *Nature Genetics*, 39(7), pp. 906–913. doi: 10.1038/ng2088.

Marees, A. T. *et al.* (2018) 'A tutorial on conducting genome-wide association studies: Quality control and statistical analysis', *International Journal of Methods in Psychiatric Research*, 27(2). doi: 10.1002/mpr.1608.

Márquez-Luna, C. *et al.* (2017) 'Multiethnic polygenic risk scores improve risk prediction in diverse populations', *Genetic Epidemiology*, 41(8), pp. 811–823. doi: 10.1002/gepi.22083.

Marshall, E. (2005) 'From dearth to deluge', *Science*, 308(5728), pp. 1570–1572. doi: 10.1126/science.308.5728.1570.

Martens, E. P. *et al.* (2006) 'Instrumental variables: Application and limitations', *Epidemiology*, 17(3), pp. 260–7. doi: 10.1097/01.ede.0000215160.88317.cb.

Martin, J. H. *et al.* (2019) 'DNA damage and repair in the female germline: Contributions to ART', *Human Reproduction Update*, 25(2), pp. 180–201. doi: 10.1093/humupd/dmy040.

Masharani, U., Goldfine, I. D. and Youngren, J. F. (2009) 'Influence of gender on the relationship between insulin sensitivity, adiposity, and plasma lipids in lean nondiabetic subjects', *Metabolism:*

clinical and experimental, 58(11), pp. 1602–8. doi: 10.1016/j.metabol.2009.05.012.

Matafome, P. and Seiça, R. (2017) 'Function and dysfunction of adipose tissue', in Letra, L. and Seiça, R. (eds) *Obesity and brain function*. Cham, pp. 5, 9–10. doi: 10.1007/978-3-319-63260-5.

Mbarek, H. *et al.* (2016) 'Identification of common genetic variants influencing spontaneous dizygotic twinning and female fertility', *American Journal of Human Genetics*, 98(5), pp. 898–908. doi: 10.1016/j.ajhg.2016.03.008.

McBryan (2013) *False discovery rates and large files*, *McBryan's Musings*. Available at: http://blog.mcbryan.co.uk/2013/01/false-discovery-rates-and-large-files.html (Accessed: 26 January 2021).

McKinney, W. (2010) 'Data structures for statistical computing in python', *Proceedings of the 9th Python in Science Conference*, pp. 51–56. doi: 10.25080/Majora-92bf1922-00a.

McMichael, A. J. *et al.* (1984) 'Dietary and endogenous cholesterol and human cancer', *Epidemiologic Reviews*, 6(1), pp. 192–216. doi: 10.1093/oxfordjournals.epirev.a036271.

Meinert, C. L. *et al.* (2000) 'Gender representation in trials', *Controlled Clinical Trials*, 21(5), pp. 462–475. doi: 10.1016/S0197-2456(00)00086-6.

Meisinger, C. *et al.* (2006) 'Body fat distribution and risk of type 2 diabetes in the general population: are there differences between men and women? The MONICA/KORA Augsburg Cohort Study', *The American Journal of Clinical Nutrition*, 84(3), pp. 483–489. doi: 10.1093/ajcn/84.3.483.

Mejia-Cristobal, L. M. *et al.* (2015) 'Tissue inhibitor of metalloproteases-4 (TIMP-4) modulates adipocyte differentiation in vitro', *Experimental Cell Research*, 335(2), pp. 207–215. doi: 10.1016/j.yexcr.2015.05.006.

Mendel, G. (1865) *Experiments in Plant Hybridization*. Available at: http://www.mendelweb.org/archive/Mendel.Experiments.txt (Accessed: 9 July 2018).

Mendis, S. *et al.* (2011) 'World Health Organization definition of myocardial infarction: 2008-09 revision', *International Journal of Epidemiology*, 40(1), pp. 139–146. doi: 10.1093/ije/dyq165.

Meyer, C., McGrath, B. P. and Teede, H. J. (2005) 'Overweight women with polycystic ovary syndrome have evidence of subclinical cardiovascular disease', *Journal of Clinical Endocrinology and Metabolism*, 90(10), pp. 5711–5716. doi: 10.1210/jc.2005-0011.

Millard, L. A. C. *et al.* (2019) 'Searching for the causal effects of body mass index in over 300 000 participants in UK Biobank, using Mendelian randomization', *PLoS Genetics*, 15(2), p. e1007951. doi: 10.1371/journal.pgen.1007951.

Mills, M. C. and Rahal, C. (2019) 'A scientometric review of genome-wide association studies', *Communications Biology*, 2(1), p. 9. doi: https://doi.org/10.1038/s42003-018-0261-x.

Misgar, R. A. *et al.* (2019) 'FSH β-subunit mutations in two sisters: the first report from the Indian sub-continent and review of previous cases', *Gynecological Endocrinology*, 35(4), pp. 290–293. doi: 10.1080/09513590.2018.1529159.

Miyashita, K. *et al.* (2009) 'Natriuretic peptides/cGMP/cGMP-dependent protein kinase cascades promote muscle mitochondrial biogenesis and prevent obesity', *Diabetes*, 58(12), pp. 2880–2892. doi: 10.2337/db09-0393.

Moran, L. J. *et al.* (2009) 'Treatment of obesity in polycystic ovary syndrome: a position statement of the androgen excess and polycystic ovary syndrome society', *Fertility and Sterility*, 92(6), pp. 1966–1982. doi: 10.1016/j.fertnstert.2008.09.018.

Moran, L. J. *et al.* (2010) 'Impaired glucose tolerance, type 2 diabetes and metabolic syndrome in polycystic ovary syndrome: a systematic review and meta-analysis', *Human Reproduction Update*, 16(4), pp. 347–363. doi: 10.1093/humupd/dmq001.

Morris, J. A. *et al.* (2019) 'An atlas of genetic influences on osteoporosis in humans and mice', *Nature Genetics*, 51(2), pp. 258–266. doi: 10.1038/s41588-018-0302-x.

Morris, R. W. *et al.* (2015) 'Heavier smoking may lead to a relative increase in waist circumference: Evidence for a causal relationship from a Mendelian randomisation meta-analysis. The CARTA consortium.', *BMJ Open*, 5(8), p. e008808. doi: 10.1136/bmjopen-2015-008808.

Morris, T. T. *et al.* (2020) 'Population phenomena inflate genetic associations of complex social traits', *Science Advances*, 6(16), p. eaay0328. doi: 10.1126/sciadv.aay0328.

Morrison, J. *et al.* (2020) 'Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics', *Nature Genetics*, 52(7), pp. 740–747. doi: 10.1038/s41588-020-0631-4.

Mukherjee, A. and Roy, S. K. (2013) 'Expression of ErbB3-binding protein-1 (EBP1) during primordial follicle formation: role of estradiol-17ß', *PLoS One*, 8(6), p. e67068. doi:

10.1371/journal.pone.0067068.

Munafò, M. R. *et al.* (2018) 'Collider scope: When selection bias can substantially influence observed associations', *International Journal of Epidemiology*, 47(1), pp. 226–235. doi: 10.1093/ije/dyx206.

National Cancer Institute at the National Institutes of Health (2020) *Breast Cancer Prevention*. Available at: https://www.cancer.gov/types/breast/patient/breast-prevention-pdq (Accessed: 24 November 2020).

National Center for Biotechnology Information (2020) *CGA glycoprotein hormones, alpha polypeptide [Homo sapiens (human)]*. Available at: https://www.ncbi.nlm.nih.gov/gene/1081 (Accessed: 25 November 2020).

National Health Service (2019) *Breast cancer in women*. Available at: https://www.nhs.uk/conditions/breast-cancer/ (Accessed: 24 November 2020).

National Institute for Health and Care Excellence (2017) *Naltrexone–bupropion for managing overweight and obesity: Technology appraisal guidance [TA494]*. Available at: https://www.nice.org.uk/guidance/ta494/chapter/1-Recommendations.

National Institute for Health and Care Excellence (2019) *Ectopic pregnancy and miscarriage: Diagnosis and initial management: NICE guideline [NG126]*. Available at: https://www.nice.org.uk/guidance/ng126 (Accessed: 26 October 2020).

National Institute for Health and Care Excellence (2020) *Tests in hospital for people with a suspected acute coronary syndrome*. Available at: https://pathways.nice.org.uk/pathways/chest-pain#path=view%3A/pathways/chest-pain/tests-in-hospital-for-people-with-a-suspected-acute-coronary-syndrome.xml&content=view-node%3Anodes-biochemical-marker-tests.

National Institute of Diabetes and Digestive and Kidney Diseases (2016) *Prescription medications to treat overweight and obesity*. Available at: https://www.niddk.nih.gov/health-information/weight-management/prescription-medications-treat-overweight-obesity (Accessed: 3 November 2020).

National Institutes of Health (2001) *NIH policy and guidelines on the inclusion of women and minorities as subjects in clinical research: NOT-OD-02-001*. Available at: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-02-001.html.

National Library of Medicine (2021) *Unified Medical Language System® (UMLS®)*. Available at: https://www.nlm.nih.gov/research/umls/index.html (Accessed: 10 February 2021).

NCD Risk Factor Collaboration (NCD-RisC) (2016a) 'A century of trends in adult human height', *eLife*, 5(e13410). doi: 10.7554/eLife.13410.

NCD Risk Factor Collaboration (NCD-RisC) (2016b) 'Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2 million participants', *Lancet*, 387(10026), pp. 1377–1396. doi: 10.1016/S0140-6736(16)30054-X.

Neale lab (2020) *UK Biobank - Neale lab*. Available at: http://www.nealelab.is/uk-biobank/ (Accessed: 4 December 2020).

Nelson, C. P. *et al.* (2017) 'Association analyses based on false discovery rate implicate new loci for coronary artery disease', *Nature Genetics*, 49(9), pp. 1385–1391. doi: 10.1038/ng.3913.

Noubiap, J. J. *et al.* (2019) 'Global prevalence of resistant hypertension: a meta-analysis of data from 3.2 million patients', *Heart*, 105(2), pp. 98–105. doi: 10.1136/heartjnl-2018-313599.

Nybacka, Å. *et al.* (2011) 'Randomized comparison of the influence of dietary management and/or physical exercise on ovarian function and metabolic parameters in overweight women with polycystic ovary syndrome', *Fertility and Sterility*, 96(6), pp. 1508–1513. doi: 10.1016/j.fertnstert.2011.09.006.

Okorodudu, D. O. *et al.* (2010) 'Diagnostic performance of body mass index to identify obesity as defined by body adiposity: a systematic review and meta-analysis', *International Journal of Obesity*, 34(5), pp. 791–799. doi: 10.1038/ijo.2010.5.

Olink (2020) *Olink's NGS-coupled protein biomarker technology to provide proteomics data from 53,000 UK Biobank participants*. Available at: https://www.olink.com/content/uploads/2020/12/Olink-UKB-PPP-Press-Release-FINAL.pdf (Accessed: 26 January 2021).

Oliphant, T. E. (2006) *A guide to NumPy*. USA: Trelgol Publishing.

Oliva, M. *et al.* (2020) 'The impact of sex on gene expression across human tissues', *Science*, 369(6509), p. eaba3066. doi: 10.1126/science.aba3066.

Olsen, M. W. B. *et al.* (2006) 'Angiopoietin-4 inhibits angiogenesis and reduces interstitial fluid

pressure', *Neoplasia*, 8(5), pp. 364–372. doi: 10.1593/neo.06127.

Ongen, H. *et al.* (2017) 'Estimating the causal tissues for complex traits and diseases', *Nature Genetics*, 49(12), pp. 1676–1683. doi: 10.1038/ng.3981.

Osborne, J. W. (2017a) 'Basic estimation and assumptions', in *Regression & linear modeling: Best practices and modern methods*. Thousand Oaks, CA: SAGE Publications, Inc, pp. 23–46. doi: 10.4135/9781071802724.

Osborne, J. W. (2017b) 'Simple linear models with categorical dependent variables: Binary logistic regression', in *Regression & linear modeling: Best practices and modern methods*. Thousand Oaks, CA: SAGE Publications, Inc, pp. 97–132. doi: 10.4135/9781071802724.

Osborne, J. W. (2017c) 'Simple linear models with continuous dependent variables: Simple regression analyses', in *Regression & linear modeling: Best practices and modern methods*. Thousand Oaks, CA: SAGE Publications, Inc, pp. 47–70. doi: 10.4135/9781071802724.

Oxenkrug, G. F. (2010) 'Tryptophan kynurenine metabolism as a common mediator of genetic and environmental impacts in major depressive disorder: the serotonin hypothesis revisited 40 years later', *The Israel Journal of Psychiatry and Related Sciences*, 47(1), pp. 56–63.

Pain, O., Dudbridge, F. and Ronald, A. (2018) 'Are your covariates under control? How normalization can re-introduce covariate effects', *European Journal of Human Genetics*, 26(8), pp. 1194–1201. doi: 10.1038/s41431-018-0159-6.

Pan, F. *et al.* (2009) 'Eos mediates Foxp3-dependent gene silencing in CD4+ regulatory T cells', *Science*, 325(5944), pp. 1142–1146. doi: 10.1126/science.1176077.

Pan, H. *et al.* (2008) 'Age-associated increase in aneuploidy and changes in gene expression in mouse eggs', *Developmental Biology*, 316(2), pp. 397–407. doi: 10.1016/j.ydbio.2008.01.048.

Pau, C. T. *et al.* (2017) 'Phenotype and tissue expression as a function of genetic risk in polycystic ovary syndrome', *PLoS One*, 12(1), p. e0168870. doi: 10.1371/journal.pone.0168870.

Pedersen, E. R. *et al.* (2015) 'Associations of plasma kynurenines with risk of acute myocardial infarction in patients with stable angina pectoris', *Atherosclerosis, Thrombosis and Vascular Biology*, 35(2), pp. 455–462. doi: 10.1161/ATVBAHA.114.304674.

Pedersen, N. L., Lichtenstein, P. and Svedberg, P. (2002) 'The swedish twin registry in the third millennium', *Twin Research*, 5(05), pp. 427–432. doi: 10.1375/twin.5.5.427.

Pereza, N. *et al.* (2017) 'Systematic review and meta-analysis of genetic association studies in idiopathic recurrent spontaneous abortion', *Fertility and Sterility*, 107(1), pp. 150-159.e2. doi: 10.1016/j.fertnstert.2016.10.007.

Perkins, M. N. and Stone, T. W. (1982) 'An iontophoretic investigation of the actions of convulsant kynurenines and their interaction with the endogenous excitant quinolinic acid', *Brain Research*, 247(1), pp. 184–187. doi: 10.1016/0006-8993(82)91048-4.

Pers, T. H. *et al.* (2015) 'Biological interpretation of genome-wide association studies using predicted gene functions', *Nature Communications*, 6, p. 5890. doi: 10.1038/ncomms6890.

Pers, T. H., Timshel, P. and Hirschhorn, J. N. (2015) 'SNPsnap: A Web-based tool for identification and annotation of matched SNPs', *Bioinformatics*, 31(3), pp. 418–420. doi: 10.1093/bioinformatics/btu655.

Persky, R. W., Turtzo, L. C. and McCullough, L. D. (2010) 'Stroke in women: Disparities and outcomes', *Current Cardiology Reports*, 12(1), pp. 6–13. doi: 10.1007/s11886-009-0080-2.

Pfaff, D., Fiedler, U. and Augustin, H. G. (2006) 'Emerging roles of the angiopoietin-tie and the ephrin-Eph systems as regulators of cell trafficking', *Journal of Leukocyte Biology*, 80(4), pp. 719–726. doi: 10.1189/jlb.1105652.

Phung, O. J. *et al.* (2014) 'Early combination therapy for the treatment of type 2 diabetes mellitus: Systematic review and meta-analysis', *Diabetes, Obesity and Metabolism*, 16(5), pp. 410–417. doi: 10.1111/dom.12233.

Pirastu, N. *et al.* (2020) 'Genetic analyses identify widespread sex-differential participation bias', *bioRxiv*. doi: 10.1101/2020.03.22.001453.

Pischon, T. *et al.* (2008) 'General and abdominal adiposity and risk of death in Europe', *New England Journal of Medicine*, 359(20), pp. 2105–2120. doi: 10.1056/NEJMoa0801891.

Pittaway, D. E., Reish, R. L. and Wentz, A. C. (1985) 'Doubling times of human chorionic gonadotropin increase in early viable intrauterine pregnancies', *American Journal of Obstetrics and Gynecology*, 152(3), pp. 299–302. doi: 10.1016/S0002-9378(85)80215-5.

Pividori, M. *et al.* (2020) 'PhenomeXcan: Mapping the genome to the phenome through the transcriptome', *Science Advances*, 6(37), p. eaba2083. doi: 10.1126/sciadv.aba2083.

Porcu, E. *et al.* (2020) 'The role of gene expression on human sexual dimorphism: Too early to

call', *bioRxiv*, p. 2020.04.15.042986. doi: 10.1101/2020.04.15.042986.

Prospective Studies Collaboration (2009) 'Body-mass index and cause-specific mortality in 900 000 adults: Collaborative analyses of 57 prospective studies', *The Lancet*, 373(9669), pp. 1083–1096. doi: 10.1016/S0140-6736(09)60318-4.

Pulit, S. L. *et al.* (2019) 'Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry', *Human Molecular Genetics*, 28(1), pp. 166–174. doi: 10.1093/hmg/ddy327.

Purcell, S. *et al.* (2007) 'PLINK: a tool set for whole-genome association and population-based linkage analyses', *The American Journal of Human Genetics*, 81(3), pp. 559–575. doi: 10.1086/519795.

Python Software Fondation (2020) *Python*. Available at: https://www.python.org/ (Accessed: 24 November 2020).

Qi, G. and Chatterjee, N. (2019) 'Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects', *Nature Communications*, 10(1), pp. 1–10. doi: 10.1038/s41467-019-09432-2.

Qiao, Y. *et al.* (2016) 'Whole exome sequencing in recurrent early pregnancy loss', *Molecular Human Reproduction*, 22(5), pp. 364–372. doi: 10.1093/molehr/gaw008.

Quintero-Ronderos, P. *et al.* (2017) 'Novel genes and mutations in patients affected by recurrent pregnancy loss', *PLoS One*, 12(10), p. e0186149. doi: 10.1371/journal.pone.0186149.

R Core Team (2014) *R: A language and environment for statistical computing*. Available at: https://www.r-project.org/ (Accessed: 24 November 2020).

Rahman, S. A. *et al.* (2019) 'Endogenous circadian regulation of female reproductive hormones', *The Journal of Clinical Endocrinology and Metabolism*, 104(12), pp. 6049–6059. doi: 10.1210/jc.2019-00803.

Ranasinghe, C. *et al.* (2013) 'Relationship between body mass index (BMI) and body fat percentage, estimated by bioelectrical impedance, in a group of Sri Lankan adults: a cross sectional study', *BMC Public Health*, 13, p. 797. doi: 10.1186/1471-2458-13-797.

Randall, J. C. *et al.* (2013) 'Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits', *PLoS Genetics*, 9(6),

p. e1003500. doi: 10.1371/journal.pgen.1003500.

Rannevik, G. *et al.* (1995) 'A longitudinal study of the perimenopausal transition: altered profiles of steroid and pituitary hormones, SHBG and bone mineral density', *Maturitas*, 21(2), pp. 103–113.

Rásky, E., Stronegger, W. J. and Freidl, W. (1996) 'The relationship between body weight and patterns of smoking in women and men', *International Journal of Epidemiology*, 25(6), pp. 1208–12. doi: 10.1093/ije/25.6.1208.

Raudvere, U. *et al.* (2019) 'g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)', *Nucleic Acids Research*, 47(W1), pp. W191–W198. doi: 10.1093/nar/gkz369.

Raulerson, C. K. *et al.* (2019) 'Adipose tissue gene expression associations reveal hundreds of candidate genes for cardiometabolic traits', *American Journal of Human Genetics*, 105(4), pp. 773–787. doi: 10.1016/j.ajhg.2019.09.001.

Rees, J. M. B. *et al.* (2019) 'Robust methods in Mendelian randomization via penalization of heterogeneous causal estimates', *PLoS Medicine*, 14(9), p. e0222362. doi: 10.1371/journal.pone.0222362.

Revelle, W. (2019) *psych: Procedures for psychological, psychometric, and personality research*. Available at: https://cran.r-project.org/package=psych (Accessed: 4 January 2021).

Ridker, P. M. *et al.* (2005) 'A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women', *New England Journal of Medicine*, 352(13). doi: 10.1056/NEJMoa050613.

Riordan, J. R. *et al.* (1989) 'Identification of the cystic fibrosis gene: Cloning and characterization of complementary DNA', *Science*, 245(4922), pp. 1066–1073. doi: 10.1126/science.2475911.

Ritchie, S. C. *et al.* (2019) 'Integrative analysis of the plasma proteome and polygenic risk of cardiometabolic diseases', *bioRxiv*, p. 2019.12.14.876474. doi: 10.1101/2019.12.14.876474.

Ritelli, M. *et al.* (2019) 'Further defining the phenotypic spectrum of B3GAT3 mutations and literature review on linkeropathy syndromes', *Genes*, 10(9), p. 631. doi: 10.3390/genes10090631.

Roadmap Epigenomics Consortium *et al.* (2015) 'Integrative analysis of 111 reference human epigenomes', *Nature*, 518(7539), pp. 317–330. doi: 10.1038/nature14248.

Rolseth, V. *et al.* (2017) 'No cancer predisposition or increased spontaneous mutation frequencies in NEIL DNA glycosylases-deficient mice', *Scientific Reports*, 7(1), p. 4384. doi: 10.1038/s41598-017-04472-4.

Romero-Corral, A. *et al.* (2008) 'Accuracy of body mass index in diagnosing obesity in the adult general population', *International Journal of Obesity*, 32(6), pp. 959–966. doi: 10.1038/ijo.2008.11.

Rommens, J. M. *et al.* (1989) 'Identification of the cystic fibrosis gene: Chromosome walking and jumping', *Science*, 245(4922), pp. 1059–1065. doi: 10.1126/science.2772657.

Roset, R. *et al.* (2014) 'The Rad50 hook domain regulates DNA damage signaling and tumorigenesis', *Genes & Development*, 28(5), pp. 451–462. doi: 10.1101/gad.236745.113.

Rost, S. *et al.* (2018) 'New indexes of body fat distribution and sex-specific risk of total and cause-specific mortality: a prospective cohort study', *BMC Public Health*, 18(1), p. 427. doi: 10.1186/s12889-018-5350-8.

Rotterdam ESHRE/ASRM-Sponsored PCOS consensus workshop group (2004) 'Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS)', *Human Reproduction*, 19(1), pp. 41–47. doi: 10.1093/humrep/deh098.

Rumbold, A. *et al.* (2011) 'Vitamin supplementation for preventing miscarriage', *Cochrane Database of Systematic Reviews*, (1). doi: 10.1002/14651858.cd004073.pub3.

Ruth, K. S. *et al.* (2020) 'Using human genetics to understand the disease impacts of testosterone in men and women', *Nature Medicine*, 26(2), pp. 252–258. doi: 10.1038/s41591-020-0751-5.

Rutten-Jacobs, L. C. A. *et al.* (2018) 'Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: Cohort study of 306 473 UK Biobank participants', *BMJ*, 363, p. 4168. doi: 10.1136/bmj.k4168.

Säfström, K., Lindahl, B. and Swahn, E. (2000) 'Risk stratification in unstable coronary artery disease', *Journal of the American College of Cardiology*, 35(7), pp. 1791–1800. doi: 10.1016/S0735-1097(00)00655-0.

Sakamuri, S. S. V. P. *et al.* (2017) 'Absence of tissue inhibitor of metalloproteinase-4 (TIMP4) ameliorates high fat diet-induced obesity in mice due to defective lipid absorption', *Scientific Reports*, 7(1), p. 6210. doi: 10.1038/s41598-017-05951-4.

Sakornsakolpat, P. *et al.* (2019) 'Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations', *Nature Genetics*, 51(3), pp. 494–505. doi: 10.1038/s41588-018-0342-2.

Sarkar, D. (2008) *Lattice: Multivariate data visualization with R*. New York: Springer.

Schreiner, P. J. *et al.* (1996) 'Sex-specific associations of magnetic resonance imaging-derived intra-abdominal and subcutaneous fat areas with conventional anthropometric indices: the atherosclerosis risk in communities study', *American Journal of Epidemiology*, 144(4), pp. 335–345. doi: 10.1093/oxfordjournals.aje.a008934.

Scuteri, A. *et al.* (2007) 'Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits', *PLoS Genetics*, 3(7), p. e115. doi: 10.1371/journal.pgen.0030115.

Seidell, J. C. *et al.* (1987) 'Assessment of intra-abdominal and subcutaneous abdominal fat: relation between anthropometry and computed tomography.', *The American Journal of Clinical Nutrition*, 45(1), pp. 7–13. doi: 10.1093/ajcn/45.1.7.

Selak, V. *et al.* (2019) 'Personalized prediction of cardiovascular benefits and bleeding harms from aspirin for primary prevention a benefit–harm analysis', *Annals of Internal Medicine*, 171(8), pp. 529–539. doi: 10.7326/M19-1132.

Sengenès, C. *et al.* (2000) 'Natriuretic peptides: a new lipolytic pathway in human adipocytes', *The FASEB Journal*, 14(10), pp. 1345–1351. doi: 10.1096/fasebj.14.10.1345.

Sherry, S. T. *et al.* (2001) 'dbSNP: the NCBI database of genetic variation', *Nucleic acids research*, 29(1), pp. 308–11. doi: 10.1093/nar/29.1.308.

Shettar, V., Patel, S. and Kidambi, S. (2017) 'Epidemiology of obesity and pharmacologic treatment options', *Nutrition in Clinical Practice*, 32(4), pp. 441–462. doi: 10.1177/0884533617713189.

Shi, Yongyong *et al.* (2012) 'Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome', *Nature Genetics*, 44(9), pp. 1020–1025. doi: 10.1038/ng.2384.

Shifman, S. *et al.* (2003) 'Linkage disequilibrium patterns of the human genome across populations', *Human Molecular Genetics*, 12(7), pp. 771–776. doi: 10.1093/hmg/ddg088.

Shu, X. *et al.* (2018) 'Associations of obesity and circulating insulin and glucose with breast cancer

risk: a Mendelian randomization analysis', *International Journal of Epidemiology*, 48(3), pp. 795–806. doi: 10.1093/ije/dyy201.

Shungin, D. *et al.* (2015) 'New genetic loci link adipose and insulin biology to body fat distribution', *Nature*, 518(7538), pp. 187–196. doi: 10.1038/nature14132.

Sierra-Johnson, J. *et al.* (2004) 'Relationships between insulin sensitivity and measures of body fat in asymptomatic men and women', *Obesity Research*, 12(12), pp. 2070–2077. doi: 10.1038/oby.2004.258.

Sivayoganathan, D. *et al.* (2011) 'Full investigation of patients with polycystic ovary syndrome (PCOS) presenting to four different clinical specialties reveals significant differences and undiagnosed morbidity', *Human Fertility*, 14(4), pp. 261–265. doi: 10.3109/14647273.2011.632058.

Smemo, S. *et al.* (2014) 'Obesity-associated variants within FTO form long-range functional connections with IRX3', *Nature*, 507(7492), pp. 371–375. doi: 10.1038/nature13138.

Sobhani, K. *et al.* (2018) 'Sex differences in ischemic heart disease and heart failure biomarkers', *Biology of Sex Differences*, 9(1), p. 43. doi: 10.1186/s13293-018-0201-y.

SomaLogic, I. (2017) *SOMAscan Proteomic Assay Technical White Paper*. Available at: https://somalogic.com/wp-content/uploads/2017/06/SSM-002-Technical-White-Paper_010916_LSM1.pdf (Accessed: 4 February 2020).

Sonestedt, E. *et al.* (2009) 'Fat and carbohydrate intake modify the association between genetic variation in the FTO genotype and obesity', *American Journal of Clinical Nutrition*, 90(5), pp. 1418–1425. doi: 10.3945/ajcn.2009.27958.

Song, P. *et al.* (2017) 'Abnormal kynurenine pathway of tryptophan catabolism in cardiovascular diseases', *Cellular and Molecular Life Sciences: CMLS*, 74(16), pp. 2899–2916. doi: 10.1007/s00018-017-2504-2.

Staiger, D. and Stock, J. H. (1997) 'Instrumental variables regression with weak instruments', *Econometrica*, 65(3), p. 557. doi: 10.2307/2171753.

van der Steeg, J. W. *et al.* (2008) 'Obesity affects spontaneous pregnancy chances in subfertile, ovulatory women', *Human Reproduction*, 23(2), pp. 324–328.

Stender, S. *et al.* (2017) 'Adiposity amplifies the genetic risk of fatty liver disease conferred by

multiple loci', *Nature genetics*, 49(6), pp. 842–847. doi: 10.1038/ng.3855.

Stone, T. W. and Darlington, L. G. (2002) 'Endogenous kynurenines as targets for drug discovery and development', *Nature Reviews. Drug Discovery*, 1(8), pp. 609–620. doi: 10.1038/nrd870.

Stone, T. W. and Perkins, M. N. (1981) 'Quinolinic acid: a potent endogenous excitant at amino acid receptors in CNS', *European Journal of Pharmacology*, 72(4), pp. 411–412. doi: 10.1016/0014-2999(81)90587-2.

Sudlow, C. *et al.* (2015) 'UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age', *PLoS Medicine*, 12(3), p. e1001779. doi: 10.1371/journal.pmed.1001779.

Sudoh, T. *et al.* (1988) 'A new natriuretic peptide in porcine brain', *Nature*, 332(6159), pp. 78–81. doi: 10.1038/332078a0.

Suhre, K. *et al.* (2017) 'Connecting genetic risk to disease end points through the human blood plasma proteome', *Nature Communications*, 8, p. 14357. doi: 10.1038/ncomms14357.

Sun, B. B. *et al.* (2018) 'Genomic atlas of the human plasma proteome.', *Nature*, 558(7708), pp. 73–79. doi: 10.1038/s41586-018-0175-2.

Swanson, S. A. and Hernán, M. A. (2013) 'Commentary: How to report instrumental variable analyses (suggestions welcome)', *Epidemiology*, 24(3), pp. 370–374. doi: 10.1097/EDE.0b013e31828d0590.

Swerdlow, D. I. *et al.* (2012) 'The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis', *The Lancet*, 379(9822), pp. 1214–1224. doi: 10.1016/S0140-6736(12)60110-X.

Swinburn, B., Sacks, G. and Ravussin, E. (2009) 'Increased food energy supply is more than sufficient to explain the US epidemic of obesity', *American Journal of Clinical Nutrition*, 90(6), pp. 1453–1456. doi: 10.3945/ajcn.2009.28595.

Szklarczyk, D. *et al.* (2019) 'STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets', *Nucleic Acids Research*, 47(D1), pp. D607–D613. doi: 10.1093/nar/gky1131.

Takeda Pharmaceuticals America, I. (2014) *Full prescribing information for CONTRAVE*. Available at: https://www.accessdata.fda.gov/drugsatfda_docs/label/2014/200063s000lbl.pdf

(Accessed: 3 November 2020).

Tarlatzis, B. C. *et al.* (2008) 'Consensus on infertility treatment related to polycystic ovary syndrome', *Human Reproduction*, 23(3), pp. 462–477. doi: 10.1093/humrep/dem426.

Taylor, A. E. *et al.* (1997) 'Determinants of abnormal gonadotropin secretion in clinically defined women with polycystic ovary syndrome', *The Journal of Clinical Endocrinology and Metabolism*, 82(7), pp. 2248–2256. doi: 10.1210/jcem.82.7.4105.

Taylor, A. E. *et al.* (2010) 'Comparison of the associations of body mass index and measures of central adiposity and fat mass with coronary heart disease, diabetes, and all-cause mortality: a study using data from 4 UK cohorts', *The American Journal of Clinical Nutrition*, 91(3), pp. 547–556. doi: 10.3945/ajcn.2009.28757.

Taylor, A. E. *et al.* (2014) 'Mendelian randomization in health research: Using appropriate genetic variants and avoiding biased estimates', *Economics and Human Biology*, 13(1), pp. 99–106. doi: 10.1016/j.ehb.2013.12.002.

Tchernof, A. and Després, J. P. (2013) 'Pathophysiology of human visceral obesity: an update', *Physiological Reviews*, 93(1), pp. 359–404. doi: 10.1152/physrev.00033.2011.

Teede, H., Deeks, A. and Moran, L. (2010) 'Polycystic ovary syndrome: a complex condition with psychological, reproductive and metabolic manifestations that impacts on health across the lifespan', *BMC Medicine*, 8, p. 41. doi: 10.1186/1741-7015-8-41.

The American College of Obstetricians and Gynecologists (2018) *Early Pregnancy Loss*, *ACOG Practice Bulletin*. Available at: https://www.acog.org/-/media/project/acog/acogorg/clinical/files/practice-bulletin/articles/2018/11/early-pregnancy-loss.pdf.

The Global Burden of Metabolic Risk Factors for Chronic Diseases Collaboration (2014) 'Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: a pooled analysis of 97 prospective cohorts with 1·8 million participants', *The Lancet*, 383(9921), pp. 970–983. doi: 10.1016/S0140-6736(13)61836-X.

The Steering Committee of the Physicians' Health Study Research (1988) 'Preliminary report: Findings from the aspirin component of the ongoing physicians' health study', *New England Journal of Medicine*, 318(4), pp. 262–264. doi: 10.1056/NEJM198801283180431.

The UK10K Consortium (2015) 'The UK10K project identifies rare variants in health and disease', *Nature*, 526(7571), pp. 82–90. doi: 10.1038/nature14962.

The World Health Organization (2000a) *Definitions and indicators in family planning maternal & child health and reproductive health used in the WHO regional office for Europe*. Available at: https://apps.who.int/iris/handle/10665/108284.

The World Health Organization (2000b) *Obesity: Preventing and Managing the Global Epidemic. Report of a WHO consultation*. Geneva: The World Health Organization.

The World Health Organization (2008) *Waist circumference and waist–hip ratio: Report of a WHO expert consultation*, *WHO*. Geneva: The World Health Organization.

The World Health Organization (2017) *Cardiovascular diseases (CVDs)*. Available at: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (Accessed: 4 February 2021).

The World Health Organization (2018a) *Global health estimates 2016: Deaths by cause, age, sex, by country and by region, 2000-2016*. Geneva: The World Health Organization.

The World Health Organization (2018b) *The top 10 causes of death*. Available at: http://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death (Accessed: 7 July 2018).

The World Health Organization (2020a) *Diabetes*. Available at: https://www.who.int/news-room/fact-sheets/detail/diabetes (Accessed: 10 November 2020).

The World Health Organization (2020b) *Obesity and overweight*. Available at: https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight (Accessed: 26 November 2020).

The World Health Organization (2020c) *The global health observatory: Body mass index*. Available at: https://www.who.int/data/gho/data/themes/theme-details/GHO/body-mass-index-(bmi) (Accessed: 26 November 2020).

Thomas, D. C., Lawlor, D. A. and Thompson, J. R. (2007) 'Re: Estimation of bias in nongenetic observational studies using "Mendelian triangulation" by Bautista et al.', *Annals of Epidemiology*, 17(7), pp. 511–513. doi: 10.1016/j.annepidem.2006.12.005.

Thomsen, S. K. *et al.* (2018) 'Type 2 diabetes risk alleles in PAM impact insulin release from

human pancreatic β-cells', *Nature Genetics*, 50(8), pp. 1122–1131. doi: 10.1038/s41588-018-0173-1.

Thrasher, J. (2017) 'Pharmacologic management of type 2 diabetes mellitus: available therapies', *American Journal of Medicine*, 130(6), pp. S4–S17. doi: 10.1016/j.amjmed.2017.04.004.

Thrift, A. P. *et al.* (2015) 'Mendelian randomization study of body mass index and colorectal cancer risk', *Cancer Epidemiology Biomarkers & Prevention*, 24(7), pp. 1024–1031. doi: 10.1158/1055-9965.EPI-14-1309.

Tobin, M. D. *et al.* (2005) 'Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure', *Statistics in Medicine*, 24(19), pp. 2911–2935. doi: 10.1002/sim.2165.

Todd, J. N. *et al.* (2015) 'Genetic evidence for a causal role of obesity in diabetic kidney disease.', *Diabetes*, 64(12), pp. 4238–46. doi: 10.2337/db15-0254.

Tola, H. *et al.* (2018) 'Assessment of the role of the anti-Mullerian hormone, luteinizing hormone/follicle stimulating hormone ratio in the diagnosis of polycystic ovary syndrome in Sudanese women', *Open Access Macedonian Journal of Medical Sciences*, 6(7), pp. 1244–1247. doi: 10.3889/oamjms.2018.260.

Trinder, J. *et al.* (2006) 'Management of miscarriage: Expectant, medical, or surgical? Results of randomised controlled trial (miscarriage treatment (MIST) trial)', *BMJ*, 332(7552), pp. 1235–1238. doi: 10.1136/bmj.38828.593125.55.

Tsuboi, N. *et al.* (2012) 'Low glomerular density with glomerulomegaly in obesity-related glomerulopathy', *Clinical journal of the American Society of Nephrology*, 7(5), pp. 735–41. doi: 10.2215/CJN.07270711.

Tsutamoto, T. *et al.* (2006) 'Relationship between renal function and plasma brain natriuretic peptide in patients with heart failure', *Journal of the American College of Cardiology*, 47(3), pp. 582–586. doi: 10.1016/j.jacc.2005.10.038.

Turcot, V. *et al.* (2018) 'Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity', *Nature Genetics*, 50(1), pp. 26–41. doi: 10.1038/s41588-017-0011-x.

U.S. Food and Drug Administration (2020) *FDA requests the withdrawal of the weight-loss drug*

Belviq, Belviq XR (lorcaserin) from the market. Available at: https://www.fda.gov/media/135189/download.

Udler, M. S. *et al.* (2018) 'Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: a soft clustering analysis', *PLoS Medicine*, 15(9), p. e1002654. doi: 10.1371/journal.pmed.1002654.

UK Biobank (2020) *UK Biobank launches one of the largest scientific studies measuring circulating proteins, to better understand the link between genetics and human disease*. Available at: https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/news/uk-biobank-launches-one-of-the-largest-scientific-studies (Accessed: 28 January 2021).

UniProt Consortium (2019) 'UniProt: a worldwide hub of protein knowledge', *Nucleic Acids Research*, 47(D1), pp. D506–D515. doi: 10.1093/nar/gky1049.

United States Department of Agriculture Economic Research Service (2020) *Food availability (per capita) data system*. Available at: https://www.ers.usda.gov/webdocs/DataFiles/50472/calories.xls?v=9400.2 (Accessed: 3 November 2020).

United States General Accounting Office (2001a) *Drug safety: Most drugs withdrawn in recent years had greater health risks for women*. Available at: https://www.gao.gov/assets/100/90642.pdf.

United States General Accounting Office (2001b) *Women sufficiently represented in new drug testing, but FDA oversight needs improvement*. Available at: https://www.gao.gov/new.items/d01754.pdf.

Utermann, G. *et al.* (1984) 'Apolipoprotein E phenotypes and hyperlipidemia', *Human Genetics*, 65(3), pp. 232–236. doi: 10.1007/BF00286508.

Valabhji, J. *et al.* (2020) 'Early outcomes from the English National Health Service diabetes prevention programme', *Diabetes Care*, 43(1), pp. 152–160. doi: 10.2337/dc19-1425.

Venter, J. C. *et al.* (2001) 'The sequence of the human genome', *Science*, 291(5507), pp. 1304–1351. doi: 10.1126/science.1058040.

Verbanck, M. *et al.* (2018) 'Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases', *Nature Genetics*, 50(5), pp. 693–698. doi: 10.1038/s41588-018-0099-7.

Verbeeten, K. C. *et al.* (2011) 'Association between childhood obesity and subsequent Type 1 diabetes: a systematic review and meta-analysis', *Diabetic Medicine*, 28(1), pp. 10–18. doi: 10.1111/j.1464-5491.2010.03160.x.

Verhaegen, J. *et al.* (2012) 'Accuracy of single progesterone test to predict early pregnancy outcome in women with pain or bleeding: Meta-analysis of cohort studies', *BMJ*, 345(7879). doi: 10.1136/bmj.e6077.

Vernon, G., Baranova, A. and Younossi, Z. M. (2011) 'Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults', *Alimentary Pharmacology & Therapeutics*, 34(3), pp. 274–285. doi: 10.1111/j.1365-2036.2011.04724.x.

Vestbo, J. *et al.* (2013) 'Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease', *American Journal of Respiratory and Critical Care Medicine*, 187(4), pp. 347–365. doi: 10.1164/rccm.201204-0596PP.

Viechtbauer, W. (2010) 'Conducting meta-analyses in R with the metafor package', *Journal of Statistical Software*, 36(3), pp. 1–48. doi: 10.18637/jss.v036.i03.

Vilsbøll, T. *et al.* (2007) 'Liraglutide, a long-acting human glucagon-like peptide-1 analog, given as monotherapy significantly improves glycemic control and lowers body weight without risk of hypoglycemia in patients with type 2 diabetes', *Diabetes Care*, 30(6), pp. 1608–1610. doi: 10.2337/dc06-2593.

Vink, J. M. *et al.* (2006) 'Heritability of polycystic ovary syndrome in a Dutch twin-family study', *Journal of Clinical Endocrinology and Metabolism*, 91(6), pp. 2100–2104. doi: 10.1210/jc.2005-1494.

Völgyi, E. *et al.* (2008) 'Assessing body composition with DXA and bioimpedance: Effects of obesity, physical activity, and age', *Obesity*, 16(3), pp. 700–705. doi: 10.1038/oby.2007.94.

Võsa, U. *et al.* (2018) 'Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis', *bioRxiv*, p. 447367. doi: 10.1101/447367.

Vozoris, N. T. and O'Donnell, D. E. (2012) 'Prevalence, risk factors, activity limitation and health care utilization of an obese, population-based sample with chronic obstructive pulmonary disease', *Canadian respiratory journal*, 19(3), pp. e18-24. doi: 10.1155/2012/732618.

de Vries, A. and Ripley, B. D. (2016) *ggdendro: create dendrograms and tree diagrams using 'ggplot2'*. Available at: https://cran.r-project.org/package=ggdendro (Accessed: 4 January 2021).

Wadden, T. A. and Frey, D. L. (1997) 'A multicenter evaluation of a proprietary weight loss program for the treatment of marked obesity: a five-year follow-up', *International Journal of Eating Disorders*, 22(2), pp. 203–212. doi: 10.1002/(SICI)1098-108X(199709)22:2<203::AID-EAT13>3.0.CO;2-1.

Waddington, C. H. (1942) 'Canalization of development and the inheritance of acquired characters', *Nature*, 150(3811), pp. 563–565. doi: 10.1038/150563a0.

Wahabi, H. A. *et al.* (2018) 'Progestogen for treating threatened miscarriage', *Cochrane Database of Systematic Reviews*, 2018(8), p. CD005943. doi: 10.1002/14651858.CD005943.pub5.

Wajchenberg, B. L. (2000) 'Subcutaneous and visceral adipose tissue: Their relation to the metabolic syndrome', *Endocrine Reviews*, pp. 697–738. doi: 10.1210/edrv.21.6.0415.

Wald, A. (1940) 'The fitting of straight lines if both variables are subject to error', *The Annals of Mathematical Statistics*, 11(3), pp. 284–300. doi: 10.1214/aoms/1177731868.

Wallace, C. (2020) 'Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses', *PLoS Genetics*, 16(4), p. e1008720. doi: 10.1371/journal.pgen.1008720.

Wallace, C. (2021) 'A more accurate method for colocalisation analysis allowing for multiple causal variants', *bioRxiv*, p. 2021.02.23.432421. doi: 10.1101/2021.02.23.432421.

Wang, D. *et al.* (2019) 'A deep proteome and transcriptome abundance atlas of 29 healthy human tissues', *Molecular Systems Biology*, 15(2). doi: 10.15252/msb.20188503.

Wang, L. *et al.* (2010) 'Genome-wide screening for risk loci of idiopathic recurrent miscarriage in a Han Chinese population: a pilot study', *Reproductive Sciences*, 17(6), pp. 578–584. doi: 10.1177/1933719110364248.

Wang, T. J. *et al.* (2004) 'Impact of obesity on plasma natriuretic peptide levels', *Circulation*, 109(5), pp. 594–600. doi: 10.1161/01.CIR.0000112582.16683.EA.

Wang, Y. *et al.* (2016) 'Estrogen and testosterone in concert with EFNB3 regulate vascular smooth muscle cell contractility and blood pressure', *American Journal of Physiology. Heart and Circulatory Physiology*, 310(7), pp. H861-72. doi: 10.1152/ajpheart.00873.2015.

Wannamethee, S. G. *et al.* (2010) 'Assessing prediction of diabetes in older adults using different adiposity measures: a 7 year prospective study in 6,923 older men and women', *Diabetologia*, 53(5), pp. 890–898. doi: 10.1007/s00125-010-1670-7.

Ward, L. D. and Kellis, M. (2012) 'HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants', *Nucleic Acids Research*, 40(D1), pp. D930–D934. doi: 10.1093/nar/gkr917.

Weeks, E. M. *et al.* (2020) 'Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases', *medRxiv*, 23, p. 28. doi: 10.1101/2020.09.08.20190561.

Wen, X. (2019) *GitHub - xqwen/integrative: Software pipeline for integrative genetic association analysis: Probabilistic assessment of enrichment and colocalization*. Available at: https://github.com/xqwen/integrative (Accessed: 2 November 2020).

Wen, X., Pique-Regi, R. and Luca, F. (2017) 'Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization', *PLoS Genetics*, 13(3), p. e1006646. doi: 10.1371/journal.pgen.1006646.

Weng, S. C. *et al.* (2018) 'Do stillbirth, miscarriage, and termination of pregnancy increase risks of attempted and completed suicide within a year? A population-based nested case–control study', *BJOG: An International Journal of Obstetrics and Gynaecology*, 125(8), pp. 983–990. doi: 10.1111/1471-0528.15105.

Wickham, H. (2011) 'The split-apply-combine strategy for data analysis', *Journal of Statistical Software*, 40(1), pp. 1–29. doi: 10.18637/jss.v040.i01.

Wickham, H. (2016) *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.

Wickham, H. *et al.* (2017) *dplyr: a grammar of data Manipulation*. Available at: https://cran.r-project.org/web/packages/dplyr/index.html (Accessed: 24 November 2020).

Wickham, H. and Henry, L. (2018) *tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. Available at: https://cran.r-project.org/package=tidyr (Accessed: 4 December 2020).

Wickham, H., Hester, J. and Chang, W. (2020) *devtools*. Available at: https://devtools.r-lib.org/ (Accessed: 4 December 2020).

Wilding, J. P. H. *et al.* (2021) 'Once-weekly semaglutide in adults with overweight or obesity', *New England Journal of Medicine*, p. NEJMoa2032183. doi: 10.1056/NEJMoa2032183.

Willer, C. J., Li, Y. and Abecasis, G. R. (2010) 'METAL: Fast and efficient meta-analysis of genomewide association scans', *Bioinformatics*, 26(17), pp. 2190–2191. doi: 10.1093/bioinformatics/btq340.

Williams, C. D. *et al.* (2011) 'Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study', *Gastroenterology*, 140(1), pp. 124–131. doi: 10.1053/J.GASTRO.2010.09.038.

Winkler, T. W. *et al.* (2015) 'EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data', *Bioinformatics*, 31(2), pp. 259–261. doi: 10.1093/bioinformatics/btu621.

Wolf, W. M. *et al.* (2018) 'The current description and future need for multidisciplinary PCOS clinics', *Journal of Clinical Medicine*, 7(11), p. 395. doi: 10.3390/jcm7110395.

Wu, H.-H. *et al.* (2003) 'Autoregulation of neurogenesis by GDF11', *Neuron*, 37(2), pp. 197–207. doi: 10.1016/s0896-6273(02)01172-8.

Yancopoulos, G. D. *et al.* (2000) 'Vascular-specific growth factors and blood vessel formation', *Nature*, 407(6801), pp. 242–248. doi: 10.1038/35025215.

Yang, J. *et al.* (2011) 'GCTA: a tool for genome-wide complex trait analysis', *The American Journal of Human Genetics*, 88(1), pp. 76–82. doi: 10.1016/j.ajhg.2010.11.011.

Yang, J. *et al.* (2012) 'Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits', *Nature Genetics*, 44(4), pp. 369–375. doi: 10.1038/ng.2213.

Yao, C. *et al.* (2017) 'Dynamic role of trans regulation of gene expression in relation to complex traits', *American Journal of Human Genetics*, 100(4), pp. 571–580. doi: 10.1016/j.ajhg.2017.02.003.

Yao, C. *et al.* (2018) 'Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease', *Nature Communications*, 9(1), pp. 1–11. doi: 10.1038/s41467-018-05512-x.

Yates, A. D. *et al.* (2020) 'Ensembl 2020', *Nucleic Acids Research*, 48(D1), pp. D682–D688. doi: 10.1093/nar/gkz966.

Yavorska, O. O. and Burgess, S. (2017) 'MendelianRandomization: an R package for performing

Mendelian randomization analyses using summarized data', *International Journal of Epidemiology*, 46(6), pp. 1734–1739. doi: 10.1093/ije/dyx034.

Yen, S. S., Vela, P. and Rankin, J. (1970) 'Inappropriate secretion of follicle-stimulating hormone and luteinizing hormone in polycystic ovarian disease', *Journal of Clinical Endocrinology and Metabolism*, 30(4), pp. 435–442. doi: 10.1210/jcem-30-4-435.

Yengo, L. *et al.* (2018) 'Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry.', *Human Molecular Genetics*, 27(20), pp. 3641–3649. doi: 10.1093/hmg/ddy271.

Yeung, E. H. *et al.* (2013) 'Adiposity and sex hormones across the menstrual cycle: the BioCycle Study', *International Journal of Obesity*, 37(2), pp. 237–243. doi: 10.1038/ijo.2012.9.

Yildiz, B. O., Knochenhauer, E. S. and Azziz, R. (2008) 'Impact of obesity on the risk for polycystic ovary syndrome', *Journal of Clinical Endocrinology and Metabolism*, 93(1), pp. 162–168. doi: 10.1210/jc.2007-1834.

Young, A. I. *et al.* (2018) 'Relatedness disequilibrium regression estimates heritability without environmental bias', *Nature Genetics*, 50(9), pp. 1304–1310. doi: 10.1038/s41588-018-0178-9.

Young, A. I. *et al.* (2019) 'Deconstructing the sources of genotype-phenotype associations in humans', *Science*, 365(6460), pp. 1396–1400. doi: 10.1126/science.aax3710.

Zawadzki, J. K. and Dunaif, A. (1992) 'Diagnostic criteria for polycystic ovary syndrome: toward a rational approach', in Dunaif, A. et al. (eds) *Polycystic ovary syndrome*. Cambridge: Blackwell Scientific Publications, pp. 377–84.

Zeng, C. *et al.* (2019) 'Liquiritin, as a natural inhibitor of AKR1C1, could interfere with the progesterone metabolism', *Frontiers in Physiology*, 10, p. 833. doi: 10.3389/fphys.2019.00833.

Zhang, T. *et al.* (2018) 'Cell-type-specific eQTL of primary melanocytes facilitates identification of melanoma susceptibility genes', *Genome research*, 28(11), pp. 1621–1635. doi: 10.1101/gr.233304.117.

Zhang, Y. *et al.* (1994) 'Positional cloning of the mouse obese gene and its human homologue', *Nature*, 372(6505), pp. 425–432. doi: 10.1038/372425a0.

Zheng, J. *et al.* (2020) 'Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases', *Nature Genetics*, 52(10), pp. 1122–1131. doi:

10.1038/s41588-020-0682-6.

Zhou, W. *et al.* (2018) ‘Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies’, *Nature Genetics*, 50(9), pp. 1335–1341. doi: 10.1038/s41588-018-0184-y.

Zhu, Z. *et al.* (2016) ‘Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets’, *Nature Genetics*, 48(5), pp. 481–487. doi: 10.1038/ng.3538.

Zuo, H. *et al.* (2016) ‘Plasma biomarkers of inflammation, the kynurenine pathway, and risks of all-cause, cancer, and cardiovascular disease mortality: the Hordaland Health Study’, *American Journal of Epidemiology*, 183(4), pp. 249–258. doi: 10.1093/aje/kwv242.

van Zuydam, N. R. *et al.* (2018) ‘A genome-wide association study of diabetic kidney disease in subjects with type 2 diabetes’, *Diabetes*, 67(7), pp. 1414–1427. doi: 10.2337/db17-0914.

Appendix

Appendix 1. Data acknowledgements, funding, publications, and data availability

Appendix 1.1. Data acknowledgements

I would like to express my gratitude to all the participants, researchers and funders that have made this work possible, especially to the UK Biobank, the Genotype-Tissue Expression project, the Genetic Investigation of ANthropometric Traits consortium, INTERVAL, and the PCOS consortium.

This research has been conducted using data from UK Biobank, a major biomedical database. This publication used data generated by the INTERVAL study. We thank the INTERVAL study volunteers, the INTERVAL study co-ordination teams (at the Universities of Cambridge and Oxford and at NHS Blood and Transplant), and blood donation staff at the 25 static centres, for their help with the INTERVAL participant recruitment and study fieldwork, as well as the Cambridge BioResource and NHSBT staff for their help with volunteer recruitment. We thank the INTERVAL Operations Team (headed by Dr Richard Houghton), the INTERVAL Data Management Team (headed by Dr Matthew Walker) and investigators involved in processing and quality control of the data (notably Ben Sun, Steve Ellis, Tao Jiang, Will Astle and Heather Elding).

Appendix 1.2. Funding

I was generously funded by an NDM Prize Studentship (reference 17/18_MSD_1108275) from the Oxford Medical Research Council Doctoral Training Partnership and the Nuffield Department of Clinical Medicine.

UK Biobank is generously supported by its founding funders the Wellcome Trust and UK Medical Research Council, as well as the Department of Health, Scottish Government, the Northwest Regional Development Agency, British Heart Foundation and Cancer Research UK. The organisation has over 150 dedicated members of staff, based in multiple locations across the UK. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The INTERVAL study is funded by NHSBT (11-01-GEN) and has been supported by the NIHR-BTRU in Donor Health and Genomics (NIHR BTRU-2014-10024) at the University of Cambridge in partnership with NHSBT. DNA extraction and genotyping was funded by the National Institute of Health Research (NIHR), the NIHR BioResource (http://bioresource.nihr.ac.uk/) and the NIHR Cambridge Biomedical Research Centre (www.cambridge-brc.org.uk). Proteomic assays were funded by the INTERVAL study academic coordinating centre and MRL, Merch & Co., Inc. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health of England, or NHSBT. The academic coordinating centre for INTERVAL was supported by core funding from: NIHR Blood and Transplant Research Unit in Donor Health and Genomics (NIHR BTRU-2014-10024), UK Medical Research Council (G0900270), British Heart Foundation (SP/09/002), and NIHR Cambridge Biomedical Research Centre. The research was supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z with additional support from the NIHR Oxford BRC. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Appendix 1.3. Publications

Publications included in the thesis:

* These authors contributed equally for first or senior authorship

Censin JC, [2 authors], Lindgren CM (2021) 'Colocalization analysis of polycystic ovary syndrome to identify potential disease-mediating genes and proteins', European Journal of Human Genetics. doi: 10.1038/s41431-021-00835-8

Laisk T*, Soares ALG*, Ferreira T*, Painter JN*, **Censin JC**, [45 authors], Granne I*, Medland SE*, Mägi R*, Lawlor DA*, Lindgren CM* (2020) 'The genetic architecture of sporadic and multiple consecutive miscarriage', Nature Communications, 11(1), p. 5980. doi: 10.1038/s41467-020-19742-5.

Censin JC, [6 authors], Holmes MV*, Lindgren CM* (2019) 'Causal relationships between obesity and the leading causes of death in women and men', PLoS Genetics, 15(10), p. e1008405. doi: 10.1371/journal.pgen.1008405.

Other publications derived from work performed during my DPhil:

* These authors contributed equally for first or senior authorship

Glastonbury CA*, Pulit SL*, Honecker J*, **Censin JC**, [12 authors], Frayling TM*, Claussnitzer M*, Lindgren CM* (2020) 'Machine Learning based histology phenotyping to investigate epidemiologic and genetic basis of adipocyte morphology and cardiometabolic traits', PLoS Computational Biology, 16(8), p. e1008044. doi: 10.1371/journal.pcbi.1008044.

Krebs K*, Bovijn J*, Zheng N*, Lepamets M, **Censin JC**, [17 authors], Holmes MV*, Lindgren CM*, Phillips EJ*, Mägi R*, Milani L*, Fadista J* (2020) 'Genome-wide study identifies association between HLA-B ∗ 55:01 and self-reported penicillin allergy', American Journal of Human Genetics, 107(4), pp. 612–621. doi: 10.1016/j.ajhg.2020.08.008.

Bovijn J, **Censin JC**, Lindgren CM, Holmes MV (2020) 'Commentary: Using human genetics to guide the repurposing of medicines', International Journal of Epidemiology, 49(4), pp. 1140–1146. doi: 10.1093/ije/dyaa015.

Bovijn J*, Krebs K*, Chen C-Y*, Boxall R*, **Censin JC**, [10 authors], Walters RG*, Mägi R*, Neale BM*, Lindgren CM*, Holmes MV* (2020) 'Evaluating the cardiovascular safety of sclerostin inhibition using evidence from meta-analysis of clinical trials and human genetics', Science Translational Medicine, 12(549), p. eaay6570. doi: 10.1126/scitranslmed.aay6570.

Bovijn J*, Jackson L*, **Censin, J***, Chen C-Y*, Laisk T*, Laber S*, [10 authors], Weedon MN, Mägi R*, Neale B*, Lindgren CM*, Murray A*, Holmes MV* (2019) 'GWAS identifies risk locus for erectile dysfunction and implicates hypothalamic neurobiology and diabetes in etiology', The American Journal of Human Genetics, 104(1), pp. 157–163. doi: 10.1016/j.ajhg.2018.11.004.

Censin JC, [2 authors], Lindgren CM (2019) 'Commentary: Mendelian randomization and women's health', International Journal of Epidemiology, 48(3), pp. 830–833. doi: 10.1093/ije/dyz141.

Thomsen SK*, Raimondo A*, Hastoy B*, [2 authors], **Censin J**, [12 authors], Gloyn AL (2018) 'Type 2 diabetes risk alleles in PAM impact insulin release from human pancreatic β-cells', Nature Genetics, 50(8), pp. 1122–1131. doi: 10.1038/s41588-018-0173-1.

Appendix 1.4. Data availability

Appendix Table 1.1. Data availability.

Data description	Webpage where data are available	Reference	Additional information
Independent LD regions	https://bitbucket.org/nygcresearch/ldetect-data/downloads/	(Berisa and Pickrell, 2016)	
European LD scores	https://github.com/bulik/ldsc	(B. K. Bulik-Sullivan *et al.*, 2015; Bulik-Sullivan and Finucane, 2017b).	
PheWAS data from the Open Targets Genetics Website	https://genetics.opentargets.org	(Carvalho-Silva et al., 2019)	
Summary statistics for PCOS	https://www.repository.cam.ac.uk/handle/1810/283491	(Day et al., 2018)	
Protein-protein interaction data	https://reactome.org/download/current/interactors/reactome.homo_sapiens.interactions.tab-delimited.txt	(Fabregat et al., 2018).	See https://reactome.org for exact level of evidence for each protein-protein interaction
Summary statistics for GTEx version 7	https://www.gtexportal.org/home/datasets	(GTEx Consortium, 2017)	
Summary statistics for REPROMETA	https://www.frontiersin.org/articles/10.3389/fgene.2019.00550/full	(Kikas et al., 2019)	
Summary statistics for FG and FI	https://www.magicinvestigators.org/downloads/	(Lagou et al., 2021)	
Summary statistics for recurrent miscarriage	http://www.geenivaramu.ee/tools/misc_sumstats.zip	(Laisk et al., 2020).	
Summary statistics for BMI, WHR, and WHRadjBMI from GIANT only	https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files	(Locke et al., 2015; Shungin et al., 2015)	
Summary statistics for SBP, DBP, and deep venous thrombosis	https://docs.google.com/spreadsheets/d/1kvPoupSzsSFBNSztMzl04xMoSC3Kcx3CrjVf4yBmESU/edit?ts=5b5f17db#gid=227859291	(Neale lab, 2020).	

Data description	Webpage where data are available	Reference	Additional information
Summary statistics for BMI, WHR, and WHRadjBMI from the GIANT and UKBB meta-analysis	https://github.com/lindgrengroup/fatdistnGWAS and https://doi.org/10.5281/zenodo.1251813	(Pulit et al., 2019)	
Rs-aliases mapping file	ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606/database/organism_data/	(Sherry et al., 2001)	The file "RsMergeArch.bcp.gz"
GRCh 37 to rs-aliases mapping file	ftp://ftp.ncbi.nih.gov/snp/organisms/archive/human_9606_b144_GRCh37p13/VCF	(Sherry et al., 2001)	The file "All_20150605.vcf.gz"
UKBB	https://www.ukbiobank.ac.uk/	(Sudlow et al., 2015)	Individual-level data available after application to researchers who meet the criteria for access
Individual-level data for INTERVAL	https://www.ebi.ac.uk/ega/home	(Sun et al., 2018)	Available through application to Sun et al. (2018) at the European Genotype Archive (accession number EGAS00001002555)
Summary-level sex-combined data for INTERVAL	http://www.phpc.cam.ac.uk/ceu/proteins/	(Sun et al., 2018)	
Summary statistics for eQTLgen	https://molgenis26.gcc.rug.nl/downloads/eqtlgen/cis-eqtl	(Võsa et al., 2018)	
HaploReg v4.1 resource	https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php	(Ward and Kellis, 2012)	
Summary statistics for miscarriage and stillbirth using SAIGE	https://www.dropbox.com/s/unomie6a8edyf4i/phenotype-information.xlsx?dl=0	(Zhou et al., 2018).	More information available at https://github.com/weizhouUMICH/SAIGE and https://www.leelabsg.org/resources
Information on EA frequencies in GTEx version 7	http://cnsgenomics.com/software/smr/#DataResource	(Zhu et al., 2016).	The file "GTEx_V7_cis_eqtl_summary.tar.gz (hg19)"

Appendix 2. Supplement to Chapter 2

Appendix Table 2.1. Case criteria.

Outcome	ICD-10	ICD-9	OPCS-4	Nurse interview, Non-cancer	Nurse interview, Cancer	Nurse interview, operation	Self-reported Diagnosis by Doctor	Description
Breast cancer	C50; C500; C501; C502; C503; C504; C505; C506; C508; C509; Z853	174; 1740; 1741; 1742; 1743; 1744; 1745; 1746; 1748; 1749; V103			1002			Codes for breast cancer, including personal history codes. Run in women only

Outcome	ICD-10	ICD-9	OPCS-4	Nurse interview, Non-cancer	Nurse interview, Cancer	Nurse interview, operation	Self-reported Diagnosis by Doctor	Description
CAD	I20; I200; I201; I208; I209; I21; I210; I211; I212; I213; I214; I219; I21X; I22; I220; I221; I228; I229; I23; I230; I231; I232; I233; I234; I235; I236; I238; I24; I240; I241; I248; I249; I251; I252; I255; I256; I258; I259	410; 4109; 411; 4119; 412; 4129; 413; 4139; 4140; 4148; 4149	K40; K401; K402; K403; K404; K408; K409; K41; K411; K412; K413; K414; K418; K419; K42; K421; K422; K423; K424; K428; K429; K43; K431; K432; K433; K434; K438; K439; K44; K441; K442; K448; K449; K45; K451; K452; K453; K454; K455; K456; K458; K459; K46; K461; K462; K463; K464; K465; K468; K469; K49; K491; K492; K493; K494; K498; K499; K501; K75; K751; K752; K753; K754; K758; K759	1074; 1075		1070; 1095; 1523	1; 2	Codes for myocardial infarction, percutaneous transluminal coronary angioplasty, coronary artery bypass grafting, chronic ischaemic heart disease and angina. Outcome definition based on Nelson et al (2017); SOFT CAD definition including angina.

Outcome	ICD-10	ICD-9	OPCS-4	Nurse interview, Non-cancer	Nurse interview, Cancer	Nurse interview, operation	Self-reported Diagnosis by Doctor	Description
CAD - control group exclusions	I250; I253; I254	4141						Participants were excluded from the CAD-control group if they had these codes pertaining to heart aneurysm and atherosclerotic cardiovascular disease. Outcome definition based on Nelson *et al.* (2017); SOFT CAD definition including angina.
COPD	J41; J410; J411; J418; J42; J43; J431; J432; J438; J439; J44; J440; J441; J448; J449	491; 4910; 4911; 4912; 4918; 4919; 492; 4929		1112; 1113; 1472			6	Codes for chronic bronchitis, emphysema, COPD, or complications specified from COPD
Chronic liver disease	K702; K703; K704; K717; K721; K74; K740; K741; K742; K743; K744; K745; K746	27103; 4562; 571; 5712; 5715; 57150; 57151; 57158; 57159; 5716		1604; 1158				Codes for fibrosis, sclerosis, cirrhosis of liver and liver failure, including if caused by alcohol, toxic liver disease, biliary cirrhosis, and other causes as well as codes defining complications specified as caused by these

Outcome	ICD-10	ICD-9	OPCS-4	Nurse interview, Non-cancer	Nurse interview, Cancer	Nurse interview, operation	Self-reported Diagnosis by Doctor	Description
Colorectal Cancer	C18; C180; C181; C182; C183; C184; C185; C186; C187; C188; C189; C19; C20	153; 1530; 1531; 1532; 1533; 1534; 1535; 1536; 1537; 1538; 1539; 154; 1540; 1541			1020; 1022; 1023			Codes for cancers from caecum to rectum, including appendix
Dementia	F00; F000; F001; F002; F009; F01; F010; F011; F012; F013; F018; F019; F03; G30; G300; G301; G308; G309	2900; 2901; 2902; 2903; 2904; 3310		1263				Codes for dementia in Alzheimer's disease, vascular dementia, unspecified dementia, senile and presenile dementia
Infertility	N97; N970; N971; N972; N973; N978; N979; N46	628; 6280; 6281; 6282; 6283; 6284; 6288; 6289; 606; 6069		1403; 1404				Codes for male or female infertility of different anatomical origins. Sex-specific codes applied in relevant sex only
Lung cancer	C33; C34; C340; C341; C342; C343; C348; C349; Z851	162; 1620; 1622; 1623; 1624; 1625; 1628; 1629; V101			1001; 1027; 1028; 1080			Codes for cancers in trachea, bronchi, and lungs, including personal history codes
NAFLD	K760							Code for fatty liver disease

Outcome	ICD-10	ICD-9	OPCS-4	Nurse interview, Non-cancer	Nurse interview, Cancer	Nurse interview, operation	Self-reported Diagnosis by Doctor	Description
Renal Failure	N17; N170; N171; N172; N178; N179; N18; N180; N181; N182; N183; N184; N185; N188; N189; N19; I120; I131; I132; Z992	584; 5845; 5846; 5847; 5848; 5849; 585; 5859; 586; 5869	L746; X40; X401; X402; X403; X404; X405; X406; X407; X408; X409; X41; X411; X412; X418; X419; X42; X421; X428; X429	1192; 1193; 1194		1476; 1580; 1581; 1582		Codes for both acute, chronic and unspecified renal failure and chronic kidney disease. Also includes renal failure from hypertensive disease and various dialysis procedures in ICD-10 and OPCS-4 codes
Renal Failure - acute	N17; N170; N171; N172; N178; N179	584; 5845; 5846; 5847; 5848; 5849						Codes that are specifically for acute renal failure
Renal Failure - chronic	N18; N180; N181; N182; N183; N184; N185; N188; N189	585; 5859						Codes that are specifically for chronic kidney disease
Renal Failure - control group exclusions	N17; N170; N171; N172; N178; N179; N18; N180; N181; N182; N183; N184; N185; N188; N189; N19; I120; I131; I132; Z992	584; 5845; 5846; 5847; 5848; 5849; 585; 5859; 586; 5869	L746; X40; X401; X402; X403; X404; X405; X406; X407; X408; X409; X41; X411; X412; X418; X419; X42; X421; X428; X429	1192; 1193; 1194		1476; 1580; 1581; 1582		Participants were excluded from the renal failure control groups, including acute renal failure and chronic kidney disease, if they had these codes pertaining to renal failure

Outcome	ICD-10	ICD-9	OPCS-4	Nurse interview, Non-cancer	Nurse interview, Cancer	Nurse interview, operation	Self-reported Diagnosis by Doctor	Description
Stroke	I60; I600; I601; I602; I603; I604; I605; I606; I607; I608; I609; I61; I610; I611; I612; I613; I614; I615; I616; I618; I619; I63; I630; I631; I632; I633; I634; I635; I636; I638; I639; I64	430; 4309; 431; 4319; 434; 4340; 4341; 4349; 436; 4369		1081; 1086; 1491; 1583			3	Codes for subarachnoid and intracerebral haemorrhages and cerebral infarctions including cerebral thrombosis and embolism, and unspecified stroke. Does not include transient cerebral ischaemia but includes acute but ill-defined cerebrovascular disease

Outcome	ICD-10	ICD-9	OPCS-4	Nurse interview, Non-cancer	Nurse interview, Cancer	Nurse interview, operation	Self-reported Diagnosis by Doctor	Description
Stroke - control group exclusions	I60; I600; I601; I602; I603; I604; I605; I606; I607; I608; I609; I61; I610; I611; I612; I613; I614; I615; I616; I618; I619; I63; I630; I631; I632; I633; I634; I635; I636; I638; I639; I64; G45; G450; G451; G452; G453; G454; G458; G459	430; 4309; 431; 4319; 434; 4340; 4341; 4349; 436; 4369; 435; 4359		1081; 1082; 1086; 1491; 1583			3	Participants were excluded from the stroke control groups, including haemorrhagic and ischaemic stroke, if they had these codes pertaining to stroke and transient ischaemic attacks

Outcome	ICD-10	ICD-9	OPCS-4	Nurse interview, Non-cancer	Nurse interview, Cancer	Nurse interview, operation	Self-reported Diagnosis by Doctor	Description
Stroke - haemorrhagic	I60; I600; I601; I602; I603; I604; I605; I606; I607; I608; I609; I61; I610; I611; I612; I613; I614; I615; I616; I618; I619	430; 4309; 431; 4319		1086; 1491				Codes that denote haemorrhagic stroke
Stroke - ischaemic	I63; I630; I631; I632; I633; I634; I635; I636; I638; I639	434; 4340; 4341; 4349		1583				Codes that denote ischaemic stroke

Outcome	ICD-10	ICD-9	OPCS-4	Nurse interview, Non-cancer	Nurse interview, Cancer	Nurse interview, operation	Self-reported Diagnosis by Doctor	Description
Type 1 diabetes								Algorithm sorts participants to likely diabetes status by using information on e.g. self-reported diabetes diagnosis, age of diagnosis, medications, start of insulin within a year of diagnosis, and other self-reported data at the baseline visit. Outcome definition from Eastwood *et al.* (2016) algorithm. Probable and possible cases or probable only. Controls defined as those denoted “diabetes unlikely”.

Outcome	ICD-10	ICD-9	OPCS-4	Nurse interview, Non-cancer	Nurse interview, Cancer	Nurse interview, operation	Self-reported Diagnosis by Doctor	Description
Type 2 diabetes								Algorithm sorts participants to likely diabetes status by using information on e.g. self-reported diabetes diagnosis, age of diagnosis, medications, start of insulin within a year of diagnosis, and other self-reported data at the baseline visit. Outcome definition from Eastwood *et al.* (2016) algorithm. Probable and possible cases or probable only. Controls defined as those denoted "diabetes unlikely".

The diagnose codes used for each code category to define cases and controls for each disease in the UKBB, or what other criteria were used to define cases and controls. The following data fields were used for each diagnosis code type; ICD-10 codes: 41202 (main diagnosis), 41204 (secondary diagnoses), 40006 (cancer register), 40001 (primary cause of death), 40002 (secondary causes of death); ICD-9 codes: 41203 (main diagnosis), 41205 (secondary diagnoses), 40013 (cancer register), OPCS-4 (operative procedures): 41200 (main code), 41210 (secondary code); Nurse interview, Non-cancer codes: 20002; Nurse interview, Cancer codes: 20001; Nurse interview, Operation codes: 20004; Self-reported diagnosis by doctor: 6150 (for stroke and CAD) and 6152 (for COPD). Table adapted (updated text and layout) from Censin, Peters, *et al.* (2019) and the corresponding pre-print (Censin, Bovijn, *et al.*, 2019).

Appendix Table 2.2. Instrument strength.

GRS	N SNPs	Sex	Trait	P	F	R^2	Estimate (95% CI)	P_{het}
BMI, combined estimates	**565**	**Combined**	**BMI**	**<1E-200**	**26,466**	**5.90%**	**1.01 (1.00,1.02)**	**-**
BMI, combined estimates	565	Men	BMI	<1E-200	12,466	6.03%	1.02 (1.00,1.04)	0.34
BMI, male sex-specific estimates	**565**	**Men**	**BMI**	**<1E-200**	**12,826**	**6.19%**	**1.02 (1.00,1.04)**	**0.04**
BMI, male P-heterogeneity Bonferroni	565	Men	BMI	<1E-200	12,490	6.04%	1.02 (1.00,1.04)	0.25
BMI, male P-heterogeneity FDR 1%	565	Men	BMI	<1E-200	12,490	6.04%	1.02 (1.00,1.04)	0.25
BMI, male P-heterogeneity FDR 5%	565	Men	BMI	<1E-200	12,539	6.06%	1.02 (1.00,1.03)	0.67
BMI, male P-heterogeneity FDR 10%	565	Men	BMI	<1E-200	12,561	6.07%	1.02 (1.00,1.03)	0.77
BMI, primary SNPs in males	220	Men	BMI	<1E-200	9,057	4.45%	1.01 (0.99,1.03)	0.05
BMI, combined estimates	565	Women	BMI	<1E-200	14,160	5.86%	1.01 (0.99,1.03)	0.34
BMI, female sex-specific estimates	**565**	**Women**	**BMI**	**<1E-200**	**14,406**	**5.96%**	**1.00 (0.98,1.01)**	**0.04**
BMI, female P-heterogeneity Bonferroni	565	Women	BMI	<1E-200	14,174	5.87%	1.01 (0.99,1.02)	0.25
BMI, female P-heterogeneity FDR 1%	565	Women	BMI	<1E-200	14,174	5.87%	1.01 (0.99,1.02)	0.25
BMI, female P-heterogeneity FDR 5%	565	Women	BMI	<1E-200	14,211	5.88%	1.01 (1.00,1.03)	0.67
BMI, female P-heterogeneity FDR 10%	565	Women	BMI	<1E-200	14,217	5.88%	1.01 (1.00,1.03)	0.77
BMI, primary SNPs in females	278	Women	BMI	<1E-200	10,265	4.32%	0.98 (0.97,1.00)	0.05
WHR, combined estimates	**324**	**Combined**	**WHR**	**<1E-200**	**14,932**	**3.41%**	**1.03 (1.02,1.05)**	**-**
WHR, combined estimates	324	Men	WHR	<1E-200	3,873	1.95%	0.78 (0.76,0.80)	1.24E-157
WHR, male sex-specific estimates	**324**	**Men**	**WHR**	**<1E-200**	**4,921**	**2.47%**	**1.01 (0.98,1.04)**	**0.04**
WHR, male P-heterogeneity Bonferroni	324	Men	WHR	<1E-200	4,636	2.33%	0.93 (0.90,0.96)	2.48E-20
WHR, male P-heterogeneity FDR 1%	324	Men	WHR	<1E-200	4,758	2.39%	0.96 (0.93,0.98)	4.74E-11
WHR, male P-heterogeneity FDR 5%	324	Men	WHR	<1E-200	4,819	2.42%	0.97 (0.94,1.00)	1.01E-07
WHR, male P-heterogeneity FDR 10%	324	Men	WHR	<1E-200	4,838	2.42%	0.98 (0.95,1.01)	8.55E-06
WHR, primary SNPs in males	77	Men	WHR	<1E-200	2,842	1.44%	0.99 (0.95,1.02)	0.02
WHR, combined estimates	324	Women	WHR	<1E-200	11,690	4.88%	1.23 (1.21,1.26)	1.24E-157

GRS	N SNPs	Sex	Trait	P	F	R^2	Estimate (95% CI)	P_{het}
WHR, female sex-specific estimates	**324**	**Women**	**WHR**	**<1E-200**	**12,661**	**5.27%**	**1.04 (1.03,1.06)**	**0.04**
WHR, female P-heterogeneity Bonferroni	324	Women	WHR	<1E-200	12,443	5.18%	1.09 (1.07,1.10)	2.48E-20
WHR, female P-heterogeneity FDR 1%	324	Women	WHR	<1E-200	12,542	5.22%	1.07 (1.05,1.09)	4.74E-11
WHR, female P-heterogeneity FDR 5%	324	Women	WHR	<1E-200	12,593	5.24%	1.06 (1.04,1.08)	1.01E-07
WHR, female P-heterogeneity FDR 10%	324	Women	WHR	<1E-200	12,604	5.24%	1.06 (1.04,1.08)	8.55E-06
WHR, primary SNPs in females	203	Women	WHR	<1E-200	11,255	4.71%	1.04 (1.02,1.06)	0.02
WHRadjBMI, combined estimates	**337**	**Combined**	**WHRadjBMI**	**<1E-200**	**19,516**	**4.42%**	**1.03 (1.02,1.05)**	**-**
WHRadjBMI, combined estimates	337	Men	WHRadjBMI	<1E-200	4,363	2.20%	0.73 (0.71,0.75)	<1E-200
WHRadjBMI, male sex-specific estimates	**337**	**Men**	**WHRadjBMI**	**<1E-200**	**5,548**	**2.78%**	**1.02 (1.00,1.05)**	**0.24**
WHRadjBMI, male P-heterogeneity Bonferroni	337	Men	WHRadjBMI	<1E-200	5,132	2.57%	0.91 (0.88,0.93)	1.04E-33
WHRadjBMI, male P-heterogeneity FDR 1%	337	Men	WHRadjBMI	<1E-200	5,311	2.66%	0.94 (0.92,0.97)	7.88E-18
WHRadjBMI, male P-heterogeneity FDR 5%	337	Men	WHRadjBMI	<1E-200	5,407	2.71%	0.97 (0.95,1.00)	1.23E-08
WHRadjBMI, male P-heterogeneity FDR 10%	337	Men	WHRadjBMI	<1E-200	5,445	2.73%	0.99 (0.97,1.02)	1.59E-04
WHRadjBMI, primary SNPs in males	90	Men	WHRadjBMI	<1E-200	3,850	1.94%	1.02 (0.98,1.05)	0.30
WHRadjBMI, combined estimates	337	Women	WHRadjBMI	<1E-200	16,431	6.74%	1.27 (1.26,1.29)	<1E-200
WHRadjBMI, female sex-specific estimates	**337**	**Women**	**WHRadjBMI**	**<1E-200**	**17,263**	**7.06%**	**1.04 (1.03,1.06)**	**0.24**
WHRadjBMI, female P-heterogeneity Bonferroni	337	Women	WHRadjBMI	<1E-200	17,026	6.97%	1.09 (1.08,1.11)	1.04E-33
WHRadjBMI, female P-heterogeneity FDR 1%	337	Women	WHRadjBMI	<1E-200	17,128	7.00%	1.08 (1.06,1.09)	7.88E-18
WHRadjBMI, female P-heterogeneity FDR 5%	337	Women	WHRadjBMI	<1E-200	17,179	7.02%	1.06 (1.05,1.08)	1.23E-08
WHRadjBMI, female P-heterogeneity FDR 10%	337	Women	WHRadjBMI	<1E-200	17,200	7.03%	1.05 (1.04,1.07)	1.59E-04
WHRadjBMI, primary SNPs in females	264	Women	WHRadjBMI	<1E-200	17,220	7.04%	1.03 (1.02,1.05)	0.30

Associations, instrument strengths (F-statistics), and trait variance explained (R^2) for the different sex-specific GRS construction approaches evaluated, computed using sex-stratified linear regression of each GRS on the relevant obesity trait after sex-stratified normalisation. All effect estimates per 1 unit higher sex-stratified GRS, which should correspond to 1 sex-stratified SD higher obesity trait. Multiple testing threshold set to P<0.002 and P_{het}<0.002, for 0.05/21 unique approaches and male-female effect estimate comparisons. The main approach used in bold. Table adapted (updated text and layout) from Censin, Peters, *et al.* (2019) and the corresponding pre-print (Censin, Bovijn, *et al.*, 2019).

Appendix Table 2.3. Sex-specific associations between GRSs for obesity and leading causes of death.

Outcome	Sex	N cases	BMI OR (95% CI)	BMI P	BMI P_{het}	WHR OR (95% CI)	WHR P	WHR P_{het}	WHRadjBMI OR (95% CI)	WHRadjBMI P	WHRadjBMI P_{het}
Type 2 Diabetes	Combined	18,329	3.08 (2.90,3.28)	4.40E-272	-	3.61 (3.32,3.93)	5.40E-199	-	2.14 (1.99,2.30)	2.80E-91	-
Type 2 Diabetes	Men	11,785	2.73 (2.53,2.95)	1.70E-142	3.40E-05	3.21 (2.85,3.63)	8.40E-80	0.10	1.90 (1.69,2.13)	1.70E-27	0.08
Type 2 Diabetes	Women	6,544	3.58 (3.23,3.96)	2.60E-134		3.68 (3.29,4.12)	4.10E-115		2.17 (1.97,2.39)	1.30E-55	
CAD	Combined	36,022	1.66 (1.59,1.74)	6.50E-102	-	1.76 (1.66,1.88)	1.80E-71	-	1.40 (1.33,1.48)	9.50E-34	-
CAD	Men	24,444	1.63 (1.54,1.73)	1.00E-64	0.63	1.73 (1.58,1.89)	8.90E-34	0.19	1.40 (1.29,1.53)	3.30E-15	0.22
CAD	Women	11,578	1.67 (1.55,1.81)	1.70E-38		1.59 (1.46,1.74)	4.70E-26		1.31 (1.22,1.41)	1.10E-12	
Breast Cancer	Women	14,325	0.98 (0.91,1.05)	0.56	-	0.95 (0.88,1.03)	0.23	-	0.97 (0.90,1.03)	0.30	-
CLD	Combined	1,367	1.64 (1.32,2.05)	1.00E-05	-	1.88 (1.40,2.53)	2.90E-05	-	1.34 (1.03,1.74)	0.03	-
CLD	Men	824	1.40 (1.05,1.85)	0.02	0.09	1.64 (1.06,2.54)	0.03	0.76	1.38 (0.91,2.11)	0.13	0.51
CLD	Women	543	2.04 (1.45,2.88)	4.70E-05		1.79 (1.22,2.63)	0.003		1.16 (0.83,1.61)	0.39	
Colorectal Cancer	Combined	5,530	1.01 (0.91,1.13)	0.83	-	1.19 (1.02,1.38)	0.02	-	1.14 (1.00,1.30)	0.05	-
Colorectal Cancer	Men	3,150	1.06 (0.91,1.22)	0.46	0.23	1.31 (1.04,1.64)	0.02	0.13	1.28 (1.03,1.60)	0.02	0.25
Colorectal Cancer	Women	2,380	0.92 (0.78,1.09)	0.34		1.05 (0.87,1.26)	0.63		1.10 (0.94,1.29)	0.25	
COPD	Combined	14,673	1.66 (1.55,1.78)	5.50E-47	-	1.47 (1.34,1.61)	6.00E-16	-	1.04 (0.96,1.13)	0.39	-
COPD	Men	7,874	1.55 (1.41,1.70)	4.90E-20	0.12	1.87 (1.62,2.17)	4.40E-17	9.20E-06	1.23 (1.07,1.42)	0.004	0.007

Outcome	Sex	N cases	BMI OR (95% CI)	BMI P	BMI P_{het}	WHR OR (95% CI)	WHR P	WHR P_{het}	WHRadjBMI OR (95% CI)	WHRadjBMI P	WHRadjBMI P_{het}
COPD	Women	6,799	1.72 (1.56,1.90)	7.60E-27		1.24 (1.11,1.38)	1.80E-04		0.98 (0.89,1.07)	0.62	
Dementia	Combined	1,030	1.36 (1.06,1.76)	0.02	-	0.73 (0.52,1.03)	0.07	-	0.68 (0.50,0.92)	0.01	-
Dementia	Men	583	1.18 (0.84,1.65)	0.34	0.13	0.58 (0.35,0.99)	0.04	0.20	0.60 (0.36,0.98)	0.04	0.30
Dementia	Women	447	1.75 (1.19,2.55)	0.004		0.91 (0.60,1.39)	0.67		0.83 (0.57,1.19)	0.31	
Infertility	Combined	1,678	1.01 (0.83,1.24)	0.89	-	1.06 (0.81,1.39)	0.67	-	1.00 (0.79,1.26)	0.97	-
Infertility	Men	85	0.68 (0.28,1.63)	0.39	0.35	2.14 (0.55,8.39)	0.27	0.3	1.37 (0.37,5.08)	0.64	0.62
Infertility	Women	1,593	1.04 (0.85,1.28)	0.67		1.03 (0.83,1.30)	0.77		0.98 (0.81,1.19)	0.86	
Lung Cancer	Combined	2,716	1.33 (1.14,1.56)	3.40E-04	-	1.16 (0.94,1.44)	0.16	-	0.94 (0.78,1.14)	0.54	-
Lung Cancer	Men	1,477	1.19 (0.96,1.47)	0.11	0.16	1.02 (0.74,1.42)	0.89	0.47	0.79 (0.57,1.08)	0.14	0.20
Lung Cancer	Women	1,239	1.48 (1.18,1.87)	7.00E-04		1.19 (0.93,1.54)	0.17		1.01 (0.81,1.26)	0.92	
NAFLD	Combined	1,684	2.92 (2.39,3.56)	5.70E-26	-	2.69 (2.06,3.51)	3.70E-13	-	1.63 (1.29,2.06)	4.80E-05	-
NAFLD	Men	909	2.71 (2.08,3.55)	2.40E-13	0.54	2.40 (1.58,3.64)	4.30E-05	0.7	1.31 (0.88,1.96)	0.18	0.41
NAFLD	Women	775	3.07 (2.31,4.10)	2.00E-14		2.66 (1.93,3.66)	2.10E-09		1.61 (1.22,2.12)	7.30E-04	
Renal Failure	Combined	9,606	1.67 (1.53,1.82)	2.70E-32	-	1.62 (1.44,1.81)	1.20E-16	-	1.16 (1.05,1.29)	0.003	-
Renal Failure	Men	5,697	1.61 (1.44,1.79)	1.60E-17	0.26	1.94 (1.63,2.30)	2.90E-14	5.30E-04	1.31 (1.11,1.54)	0.001	0.05
Renal Failure	Women	3,909	1.77 (1.56,2.02)	6.10E-18		1.31 (1.13,1.51)	3.00E-04		1.06 (0.94,1.20)	0.34	
Renal Failure - Acute	Combined	4,686	1.82 (1.61,2.05)	2.70E-22	-	1.57 (1.33,1.84)	5.20E-08	-	1.12 (0.97,1.29)	0.12	-

Outcome	Sex	N cases	BMI OR (95% CI)	BMI P	BMI P_{het}	WHR OR (95% CI)	WHR P	WHR P_{het}	WHRadjBMI OR (95% CI)	WHRadjBMI P	WHRadjBMI P_{het}
Renal Failure - Acute	Men	3,035	1.73 (1.49,2.01)	3.40E-13	0.32	1.87 (1.48,2.35)	1.20E-07	0.01	1.18 (0.94,1.47)	0.15	0.54
Renal Failure - Acute	Women	1,651	1.97 (1.61,2.40)	2.30E-11		1.25 (1.00,1.56)	0.05		1.07 (0.89,1.30)	0.47	
Renal Failure - Chronic	Combined	4,599	1.81 (1.60,2.05)	9.90E-22	-	1.74 (1.48,2.05)	2.50E-11	-	1.22 (1.06,1.41)	0.006	-
Renal Failure - Chronic	Men	2,581	1.87 (1.59,2.19)	2.50E-14	0.71	2.33 (1.82,3.00)	3.40E-11	1.30E-04	1.67 (1.31,2.12)	3.00E-05	3.50E-04
Renal Failure - Chronic	Women	2,018	1.78 (1.49,2.13)	3.00E-10		1.25 (1.02,1.53)	0.03		0.97 (0.82,1.16)	0.75	
Stroke	Combined	10,788	1.40 (1.30,1.52)	1.00E-16	-	1.35 (1.21,1.50)	3.70E-08	-	1.15 (1.04,1.26)	0.005	-
Stroke	Men	6,334	1.42 (1.28,1.57)	3.80E-11	0.84	1.57 (1.33,1.84)	5.80E-08	0.009	1.22 (1.05,1.42)	0.01	0.40
Stroke	Women	4,454	1.39 (1.23,1.57)	8.60E-08		1.18 (1.03,1.35)	0.02		1.12 (1.00,1.26)	0.05	
Stroke - Haemorrhagic	Combined	1,907	1.25 (1.03,1.50)	0.02	-	1.37 (1.07,1.76)	0.01	-	1.15 (0.92,1.44)	0.2	-
Stroke - Haemorrhagic	Men	928	1.24 (0.95,1.62)	0.11	0.90	1.38 (0.91,2.09)	0.13	0.54	1.11 (0.74,1.64)	0.62	0.99
Stroke - Haemorrhagic	Women	979	1.27 (0.98,1.64)	0.07		1.18 (0.89,1.57)	0.25		1.10 (0.86,1.41)	0.44	
Stroke - Ischaemic	Combined	3,351	1.38 (1.20,1.59)	8.10E-06	-	1.20 (1.00,1.46)	0.06	-	1.16 (0.98,1.37)	0.09	-
Stroke - Ischaemic	Men	2,171	1.43 (1.20,1.70)	6.50E-05	0.56	1.33 (1.01,1.74)	0.04	0.33	1.27 (0.98,1.65)	0.07	0.72
Stroke - Ischaemic	Women	1,180	1.31 (1.04,1.65)	0.02		1.10 (0.85,1.43)	0.47		1.19 (0.95,1.49)	0.12	
Type 1 Diabetes	Combined	1,503	1.65 (1.34,2.04)	3.10E-06	-	1.15 (0.86,1.52)	0.35	-	1.06 (0.83,1.36)	0.64	-
Type 1 Diabetes	Men	826	1.46 (1.10,1.93)	0.009	0.23	1.15 (0.74,1.78)	0.54	0.81	1.29 (0.85,1.97)	0.23	0.22
Type 1 Diabetes	Women	677	1.89 (1.39,2.57)	5.40E-05		1.07 (0.76,1.51)	0.69		0.94 (0.70,1.26)	0.67	

Effect estimates computed using sex-stratified logistic regression between the GRS and each disease, with effect estimates per 1 sex-stratified higher GRS. Multiple testing thresholds set to P<0.001, for 0.05/51 unique obesity trait-disease combinations in the study and P_{het}<0.001, for 0.05/48 unique male-female comparisons between the obesity trait-disease effect estimates (fewer comparisons since breast cancer was investigated in women only). Table adapted (updated text and layout) from Censin, Peters, *et al.* (2019) and the corresponding pre-print (Censin, Bovijn, *et al.*, 2019).

Appendix Table 2.4. Sex-specific effects of obesity on blood pressure traits, computed using MR and using different approaches to construct the sex-specific GRSs.

Exposure	Outcome	GRS	Sex	Estimate, Clinical (95% CI)	P (Clinical)	P_{het} (Clinical)	Estimate, SD (95% CI)	P (SD)	P_{het} (SD)
BMI	**DBP**	**BMI, combined estimates**	**Combined**	**2.92 (2.78,3.07)**	**<1E-200**	**-**	**0.27 (0.26,0.28)**	**<1E-200**	**-**
BMI	DBP	BMI, combined GIANT weights	Combined	2.67 (2.51,2.83)	<1E-200	-	0.25 (0.23,0.26)	<1E-200	-
BMI	DBP	BMI, unweighted	Combined	3.09 (2.94,3.24)	<1E-200	-	0.29 (0.27,0.30)	<1E-200	-
BMI	**DBP**	**BMI, male sex-specific estimates**	**Men**	**2.60 (2.39,2.80)**	**1.30E-136**	**1.83E-05**	**0.24 (0.22,0.26)**	**1.58E-140**	**5.24E-05**
BMI	DBP	BMI, male GIANT weights	Men	2.31 (2.07,2.54)	1.01E-81	6.79E-05	0.22 (0.19,0.24)	2.10E-84	2.08E-04
BMI	DBP	BMI, unweighted	Men	2.77 (2.54,2.99)	8.37E-128	2.40E-04	0.26 (0.24,0.28)	1.50E-131	5.62E-04
BMI	**DBP**	**BMI, female sex-specific estimates**	**Women**	**3.22 (3.02,3.41)**	**<1E-200**	**1.83E-05**	**0.30 (0.28,0.31)**	**<1E-200**	**5.24E-05**
BMI	DBP	BMI, female GIANT weights	Women	2.97 (2.74,3.19)	4.40E-149	6.79E-05	0.27 (0.25,0.29)	6.15E-149	2.08E-04
BMI	DBP	BMI, unweighted	Women	3.35 (3.13,3.56)	<1E-200	2.40E-04	0.31 (0.29,0.33)	<1E-200	5.62E-04
BMI	**SBP**	**BMI, combined estimates**	**Combined**	**3.52 (3.28,3.77)**	**5.67E-176**	**-**	**0.19 (0.18,0.20)**	**1.52E-184**	**-**
BMI	SBP	BMI, combined GIANT weights	Combined	3.14 (2.87,3.42)	3.94E-113	-	0.17 (0.16,0.18)	4.78E-118	-
BMI	SBP	BMI, unweighted	Combined	3.73 (3.47,3.99)	2.59E-170	-	0.20 (0.19,0.22)	8.32E-179	-
BMI	**SBP**	**BMI, male sex-specific estimates**	**Men**	**3.41 (3.07,3.75)**	**1.13E-85**	**0.42**	**0.19 (0.17,0.21)**	**2.03E-90**	**0.99**
BMI	SBP	BMI, male GIANT weights	Men	2.99 (2.59,3.38)	4.14E-50	0.38	0.17 (0.15,0.19)	1.26E-52	0.82
BMI	SBP	BMI, unweighted	Men	3.74 (3.36,4.11)	5.42E-85	0.84	0.21 (0.19,0.23)	2.05E-89	0.32
BMI	**SBP**	**BMI, female sex-specific estimates**	**Women**	**3.61 (3.27,3.95)**	**4.07E-97**	**0.42**	**0.19 (0.17,0.21)**	**5.69E-101**	**0.99**
BMI	SBP	BMI, female GIANT weights	Women	3.23 (2.85,3.62)	2.87E-60	0.38	0.17 (0.15,0.19)	3.58E-62	0.82
BMI	SBP	BMI, unweighted	Women	3.69 (3.32,4.05)	8.62E-88	0.84	0.20 (0.18,0.21)	2.73E-91	0.32
WHR	**DBP**	**WHR, combined estimates**	**Combined**	**2.68 (2.49,2.86)**	**1.40E-171**	**-**	**0.25 (0.23,0.26)**	**6.13E-174**	**-**
WHR	DBP	WHR, combined GIANT weights	Combined	2.54 (2.32,2.75)	4.44E-119	-	0.23 (0.21,0.25)	2.24E-120	-

Exposure	Outcome	GRS	Sex	Estimate, Clinical (95% CI)	P (Clinical)	P_{het} (Clinical)	Estimate, SD (95% CI)	P (SD)	P_{het} (SD)
WHR	DBP	WHR, unweighted	Combined	2.83 (2.62,3.03)	3.17E-163	-	0.26 (0.24,0.28)	8.30E-166	-
WHR	**DBP**	**WHR, male sex-specific estimates**	**Men**	**2.80 (2.47,3.12)**	**2.53E-63**	**0.01**	**0.26 (0.23,0.29)**	**1.53E-64**	**0.01**
WHR	DBP	WHR, male GIANT weights	Men	2.84 (2.44,3.24)	2.03E-43	0.08	0.26 (0.23,0.30)	2.98E-44	0.07
WHR	DBP	WHR, unweighted	Men	2.64 (2.27,3.01)	5.69E-44	0.98	0.25 (0.21,0.28)	7.48E-45	0.96
WHR	**DBP**	**WHR, female sex-specific estimates**	**Women**	**2.30 (2.10,2.51)**	**3.95E-108**	**0.01**	**0.21 (0.19,0.23)**	**8.97E-110**	**0.01**
WHR	DBP	WHR, female GIANT weights	Women	2.42 (2.17,2.67)	8.79E-81	0.08	0.22 (0.20,0.25)	1.19E-81	0.07
WHR	DBP	WHR, unweighted	Women	2.63 (2.40,2.87)	4.21E-109	0.98	0.24 (0.22,0.27)	4.86E-111	0.96
WHR	**SBP**	**WHR, combined estimates**	**Combined**	**3.98 (3.66,4.30)**	**2.81E-129**	**-**	**0.22 (0.20,0.23)**	**2.56E-134**	**-**
WHR	SBP	WHR, combined GIANT weights	Combined	3.89 (3.52,4.26)	7.08E-95	-	0.21 (0.19,0.23)	8.26E-99	-
WHR	SBP	WHR, unweighted	Combined	4.33 (3.98,4.68)	5.60E-130	-	0.23 (0.22,0.25)	1.32E-134	-
WHR	**SBP**	**WHR, male sex-specific estimates**	**Men**	**3.61 (3.07,4.15)**	**4.45E-39**	**0.94**	**0.20 (0.17,0.23)**	**1.63E-40**	**0.55**
WHR	SBP	WHR, male GIANT weights	Men	3.96 (3.29,4.63)	4.71E-31	0.93	0.22 (0.18,0.26)	3.80E-32	0.61
WHR	SBP	WHR, unweighted	Men	3.84 (3.22,4.46)	7.63E-34	0.65	0.21 (0.18,0.25)	6.75E-35	0.99
WHR	**SBP**	**WHR, female sex-specific estimates**	**Women**	**3.59 (3.23,3.95)**	**2.43E-85**	**0.94**	**0.19 (0.17,0.21)**	**5.02E-89**	**0.55**
WHR	SBP	WHR, female GIANT weights	Women	3.93 (3.49,4.36)	5.35E-69	0.93	0.21 (0.19,0.23)	2.16E-72	0.61
WHR	SBP	WHR, unweighted	Women	4.01 (3.61,4.42)	7.00E-83	0.65	0.21 (0.19,0.23)	2.72E-86	0.99
WHRadjBMI	**DBP**	**WHRadjBMI, combined estimates**	**Combined**	**1.07 (0.91,1.23)**	**1.87E-38**	**-**	**0.10 (0.08,0.11)**	**3.24E-39**	**-**
WHRadjBMI	DBP	WHRadjBMI, combined GIANT weights	Combined	0.92 (0.73,1.10)	1.60E-22	-	0.09 (0.07,0.10)	4.57E-23	-
WHRadjBMI	DBP	WHRadjBMI, unweighted	Combined	0.86 (0.68,1.04)	1.29E-20	-	0.08 (0.06,0.10)	2.66E-21	-
WHRadjBMI	**DBP**	**WHRadjBMI, male sex-specific estimates**	**Men**	**0.96 (0.66,1.26)**	**3.50E-10**	**0.75**	**0.09 (0.06,0.12)**	**2.51E-10**	**0.74**

Exposure	Outcome	GRS	Sex	Estimate, Clinical (95% CI)	P (Clinical)	P_{het} (Clinical)	Estimate, SD (95% CI)	P (SD)	P_{het} (SD)
WHRadjBMI	DBP	WHRadjBMI, male GIANT weights	Men	1.21 (0.81,1.61)	2.06E-09	0.36	0.11 (0.07,0.15)	1.93E-09	0.39
WHRadjBMI	DBP	WHRadjBMI, unweighted	Men	0.77 (0.41,1.13)	2.26E-05	0.90	0.07 (0.04,0.10)	1.67E-05	0.89
WHRadjBMI	**DBP**	**WHRadjBMI, female sex-specific estimates**	**Women**	**1.02 (0.84,1.19)**	**2.36E-30**	**0.75**	**0.09 (0.08,0.11)**	**6.41E-31**	**0.74**
WHRadjBMI	DBP	WHRadjBMI, female GIANT weights	Women	1.00 (0.79,1.21)	5.14E-20	0.36	0.09 (0.07,0.11)	1.63E-20	0.39
WHRadjBMI	DBP	WHRadjBMI, unweighted	Women	0.80 (0.59,1.00)	8.91E-15	0.90	0.07 (0.06,0.09)	2.65E-15	0.89
WHRadjBMI	**SBP**	**WHRadjBMI, combined estimates**	**Combined**	**2.27 (1.99,2.55)**	**7.50E-57**	**-**	**0.12 (0.11,0.14)**	**2.97E-59**	**-**
WHRadjBMI	SBP	WHRadjBMI, combined GIANT weights	Combined	2.13 (1.81,2.45)	3.69E-39	-	0.12 (0.10,0.13)	7.48E-41	-
WHRadjBMI	SBP	WHRadjBMI, unweighted	Combined	2.25 (1.94,2.56)	3.43E-45	-	0.12 (0.11,0.14)	3.21E-47	-
WHRadjBMI	**SBP**	**WHRadjBMI, male sex-specific estimates**	**Men**	**1.58 (1.07,2.08)**	**8.41E-10**	**0.04**	**0.09 (0.06,0.12)**	**4.38E-10**	**0.08**
WHRadjBMI	SBP	WHRadjBMI, male GIANT weights	Men	2.33 (1.67,3.00)	6.20E-12	1.00	0.13 (0.09,0.17)	3.59E-12	0.86
WHRadjBMI	SBP	WHRadjBMI, unweighted	Men	2.12 (1.52,2.72)	3.87E-12	0.84	0.12 (0.09,0.15)	1.62E-12	0.65
WHRadjBMI	**SBP**	**WHRadjBMI, female sex-specific estimates**	**Women**	**2.18 (1.87,2.49)**	**7.24E-44**	**0.04**	**0.12 (0.10,0.13)**	**3.69E-46**	**0.08**
WHRadjBMI	SBP	WHRadjBMI, female GIANT weights	Women	2.33 (1.96,2.71)	1.33E-33	1.00	0.13 (0.11,0.14)	1.30E-35	0.86
WHRadjBMI	SBP	WHRadjBMI, unweighted	Women	2.05 (1.70,2.41)	1.14E-29	0.84	0.11 (0.09,0.13)	3.81E-31	0.65

MR estimates computed using the ratio method. All effect estimates per 1 sex-stratified SD unit higher obesity trait, with the outcome in either clinical or sex-stratified SD units, as specified. Multiple testing thresholds set to $P<0.003$ and $P_{het}<0.003$, for 0.05/15 unique obesity trait-risk factors evaluated and male-female effect estimate comparisons. Table adapted (updated text and layout) from Censin, Peters, *et al.* (2019) and the corresponding pre-print (Censin, Bovijn, *et al.*, 2019).

Appendix Table 2.5. Sex-specific effects of obesity on smoking status, computed using MR and using different GRS weighting approaches, with analyses performed in both Europeans and the British ancestry only subset.

Exposure	Outcome	GRS	Sex	OR (95% CI), Europeans	P (Europeans)	P_{het} (Europeans)	OR (95% CI), British	P (British)	P_{het} (British)
BMI	Having smoked/smoker	BMI, combined estimates	Combined	1.32 (1.29, 1.35)	4.20E-100	-	1.32 (1.28, 1.35)	4.27E-89	-
BMI	Having smoked/smoker	BMI, combined GIANT weights	Combined	1.28 (1.25, 1.32)	2.18E-64	-	1.28 (1.24, 1.31)	3.66E-55	-
BMI	Having smoked/smoker	BMI, unweighted	Combined	1.32 (1.29, 1.36)	2.89E-89	-	1.32 (1.28, 1.36)	1.40E-79	-
BMI	Having smoked/smoker	BMI, male sex-specific estimates	Men	1.37 (1.32, 1.43)	2.20E-63	8.38E-04	1.36 (1.31, 1.42)	1.87E-54	0.01
BMI	Having smoked/smoker	BMI, male GIANT weights	Men	1.32 (1.27, 1.38)	1.19E-37	0.05	1.30 (1.25, 1.36)	8.29E-31	0.19
BMI	Having smoked/smoker	BMI, unweighted	Men	1.41 (1.36, 1.47)	3.52E-62	2.31E-05	1.40 (1.34, 1.46)	2.61E-54	4.61E-04
BMI	Having smoked/smoker	BMI, female sex-specific estimates	Women	1.26 (1.22, 1.30)	1.21E-38	8.38E-04	1.27 (1.22, 1.32)	4.19E-36	0.01
BMI	Having smoked/smoker	BMI, female GIANT weights	Women	1.25 (1.20, 1.30)	5.61E-27	0.05	1.25 (1.20, 1.31)	3.86E-24	0.19
BMI	Having smoked/smoker	BMI, unweighted	Women	1.25 (1.21, 1.30)	1.70E-32	2.31E-05	1.26 (1.21, 1.31)	4.59E-30	4.61E-04
WHR	Having smoked/smoker	WHR, combined estimates	Combined	1.23 (1.19, 1.27)	2.20E-34	-	1.22 (1.18, 1.26)	1.86E-29	-
WHR	Having smoked/smoker	WHR, combined GIANT weights	Combined	1.26 (1.21, 1.31)	4.69E-33	-	1.25 (1.20, 1.30)	4.73E-28	-
WHR	Having smoked/smoker	WHR, unweighted	Combined	1.25 (1.21, 1.30)	6.44E-35	-	1.25 (1.20, 1.30)	1.03E-30	-
WHR	Having smoked/smoker	WHR, male sex-specific estimates	Men	1.47 (1.38, 1.56)	1.06E-35	2.24E-14	1.44 (1.35, 1.53)	1.15E-29	2.41E-11
WHR	Having smoked/smoker	WHR, male GIANT weights	Men	1.49 (1.38, 1.60)	4.85E-26	3.27E-09	1.47 (1.36, 1.58)	4.52E-22	1.38E-07
WHR	Having smoked/smoker	WHR, unweighted	Men	1.45 (1.35, 1.55)	5.58E-26	6.95E-08	1.43 (1.33, 1.53)	3.71E-22	4.60E-06

Exposure	Outcome	GRS	Sex	OR (95% CI), Europeans	P (Europeans)	P_{het} (Europeans)	OR (95% CI), British	P (British)	P_{het} (British)
WHR	Having smoked/smoker	WHR, female sex-specific estimates	Women	1.12 (1.08, 1.16)	3.35E-09	2.24E-14	1.12 (1.08, 1.16)	1.12E-08	2.41E-11
WHR	Having smoked/smoker	WHR, female GIANT weights	Women	1.15 (1.10, 1.20)	8.37E-10	3.27E-09	1.15 (1.10, 1.20)	8.78E-09	1.38E-07
WHR	Having smoked/smoker	WHR, unweighted	Women	1.16 (1.11, 1.21)	1.47E-12	6.95E-08	1.17 (1.12, 1.22)	2.94E-12	4.60E-06
WHRadjBMI	Having smoked/smoker	WHRadjBMI, combined estimates	Combined	1.03 (1.00, 1.06)	0.07	-	1.03 (1.00, 1.06)	0.07	-
WHRadjBMI	Having smoked/smoker	WHRadjBMI, combined GIANT weights	Combined	1.00 (0.97, 1.03)	0.89	-	1.00 (0.96, 1.03)	0.87	-
WHRadjBMI	Having smoked/smoker	WHRadjBMI, unweighted	Combined	1.02 (0.99, 1.05)	0.26	-	1.02 (0.99, 1.06)	0.23	-
WHRadjBMI	Having smoked/smoker	WHRadjBMI, male sex-specific estimates	Men	1.10 (1.04, 1.16)	0.001	0.006	1.10 (1.03, 1.16)	0.002	0.01
WHRadjBMI	Having smoked/smoker	WHRadjBMI, male GIANT weights	Men	1.07 (1.00, 1.16)	0.06	0.05	1.08 (0.99, 1.17)	0.07	0.06
WHRadjBMI	Having smoked/smoker	WHRadjBMI, unweighted	Men	1.11 (1.04, 1.18)	0.003	0.005	1.11 (1.04, 1.20)	0.003	0.007
WHRadjBMI	Having smoked/smoker	WHRadjBMI, female sex-specific estimates	Women	1.00 (0.97, 1.03)	0.85	0.006	1.01 (0.97, 1.04)	0.71	0.01
WHRadjBMI	Having smoked/smoker	WHRadjBMI, female GIANT weights	Women	0.99 (0.95, 1.03)	0.53	0.05	0.99 (0.95, 1.03)	0.58	0.06
WHRadjBMI	Having smoked/smoker	WHRadjBMI, unweighted	Women	0.99 (0.96, 1.03)	0.74	0.005	1.00 (0.96, 1.04)	0.88	0.007

MR estimates computed using the ratio method. All effect estimates per 1 sex-stratified SD unit higher obesity trait. Multiple testing thresholds set to $P<0.003$ and $P_{het}<0.003$, for 0.05/15 unique obesity trait-risk factors evaluated and male-female effect estimate comparisons. Table adapted (updated text and layout) from Censin, Peters, *et al.* (2019) and the corresponding pre-print (Censin, Bovijn, *et al.*, 2019).

Appendix Table 2.6. Sex-specific effects of obesity on blood pressure traits in clinical and SD units, conducted in the British ancestry subset.

Exposure	Outcome	Sex	Estimate, Clinical (95% CI)	P (Clinical)	P_{het} (Clinical)	Estimate, SD (95% CI)	P (SD)	P_{het} (SD)
BMI	DBP	Combined	2.91 (2.76,3.06)	<1E-200	-	0.27 (0.26,0.28)	<1E-200	-
BMI	DBP	Men	2.59 (2.38,2.81)	2.44E-123	4.63E-05	0.24 (0.22,0.26)	4.88E-127	1.24E-04
BMI	DBP	Women	3.21 (3.01,3.42)	3.18E-200		0.30 (0.28,0.31)	<1E-200	
BMI	SBP	Combined	3.51 (3.25,3.77)	4.26E-155	-	0.19 (0.18,0.20)	8.29E-163	-
BMI	SBP	Men	3.40 (3.04,3.75)	8.05E-77	0.45	0.19 (0.17,0.21)	3.36E-81	0.97
BMI	SBP	Women	3.59 (3.23,3.95)	1.69E-84		0.19 (0.17,0.21)	8.35E-88	
WHR	DBP	Combined	2.66 (2.46,2.86)	2.45E-152	-	0.25 (0.23,0.26)	2.79E-154	-
WHR	DBP	Men	2.82 (2.48,3.16)	1.56E-58	0.009	0.26 (0.23,0.29)	8.56E-60	0.007
WHR	DBP	Women	2.28 (2.06,2.49)	5.07E-94		0.21 (0.19,0.23)	3.00E-95	
WHR	SBP	Combined	3.99 (3.65,4.33)	5.35E-116	-	0.22 (0.20,0.23)	1.76E-120	-
WHR	SBP	Men	3.64 (3.07,4.21)	4.84E-36	0.87	0.20 (0.17,0.23)	2.31E-37	0.52
WHR	SBP	Women	3.59 (3.21,3.97)	1.43E-75		0.19 (0.17,0.21)	5.63E-79	
WHRadjBMI	DBP	Combined	1.06 (0.89,1.23)	2.07E-34	-	0.10 (0.08,0.11)	6.77E-35	-
WHRadjBMI	DBP	Men	0.98 (0.66,1.30)	1.20E-09	0.94	0.09 (0.06,0.12)	9.88E-10	0.94
WHRadjBMI	DBP	Women	0.99 (0.81,1.18)	2.85E-26		0.09 (0.07,0.11)	1.19E-26	
WHRadjBMI	SBP	Combined	2.27 (1.97,2.57)	2.76E-51	-	0.12 (0.11,0.14)	2.63E-53	-
WHRadjBMI	SBP	Men	1.62 (1.09,2.15)	2.28E-09	0.09	0.09 (0.06,0.12)	1.30E-09	0.14
WHRadjBMI	SBP	Women	2.16 (1.84,2.49)	6.30E-39		0.12 (0.10,0.13)	5.93E-41	

MR estimates computed using the ratio method. All effect estimates per 1 sex-stratified SD unit higher obesity trait, with the outcome in either clinical units (mmHg) or sex-stratified SD units, as specified. Multiple testing thresholds set to P<0.003 and P_{het}<0.003, for 0.05/15 unique obesity trait-risk factors evaluated and male-female effect estimate comparisons. Table adapted (updated text and layout) from Censin, Peters, *et al.* (2019) and the corresponding pre-print (Censin, Bovijn, *et al.*, 2019).

Appendix Table 2.7. Logistic regression results of association testing between the obesity genetic risk score and disease outcomes using the same number of cases and controls in men and women.

Outcome	GRS	N SNPs	Sex	OR (95% CI), all available cases and controls for men and women	P, all available cases and controls for men and women	P_{het}, all available cases and controls for men and women	OR (95% CI), same number of cases and controls for men and women	P, same number of cases and controls for men and women	P_{het}, same number of cases and controls for men and women
CAD	BMI, men	565	Men	1.63 (1.54, 1.73)	9.97E-65	0.63	1.66 (1.54, 1.80)	8.73E-37	0.81
CAD	BMI, women	565	Women	1.67 (1.55, 1.81)	1.70E-38		1.69 (1.56, 1.82)	8.24E-39	
CAD	WHR, men	324	Men	1.73 (1.58, 1.89)	8.94E-34	0.19	1.76 (1.55, 1.99)	2.50E-19	0.23
CAD	WHR, women	324	Women	1.59 (1.46, 1.74)	4.67E-26		1.60 (1.47, 1.75)	4.44E-26	
CAD	WHRadjBMI, men	337	Men	1.40 (1.29, 1.53)	3.31E-15	0.22	1.37 (1.22, 1.54)	1.57E-07	0.52
CAD	WHRadjBMI, women	337	Women	1.31 (1.22, 1.41)	1.15E-12		1.31 (1.21, 1.41)	2.65E-12	
CLD	BMI, men	565	Men	1.40 (1.05, 1.85)	0.02	0.09	1.39 (0.98, 1.96)	0.06	0.12
CLD	BMI, women	565	Women	2.04 (1.45, 2.88)	4.71E-05		2.04 (1.45, 2.88)	4.74E-05	
CLD	WHR, men	324	Men	1.64 (1.06, 2.54)	0.03	0.76	1.68 (0.98, 2.88)	0.06	0.85
CLD	WHR, women	324	Women	1.79 (1.22, 2.63)	0.003		1.79 (1.22, 2.62)	0.003	
CLD	WHRadjBMI, men	337	Men	1.38 (0.91, 2.11)	0.13	0.51	1.73 (1.03, 2.91)	0.04	0.19
CLD	WHRadjBMI, women	337	Women	1.16 (0.83, 1.61)	0.39		1.15 (0.83, 1.61)	0.40	
COPD	BMI, men	565	Men	1.55 (1.41, 1.70)	4.92E-20	0.12	1.57 (1.42, 1.73)	1.96E-18	0.17
COPD	BMI, women	565	Women	1.72 (1.56, 1.90)	7.59E-27		1.73 (1.56, 1.91)	7.23E-27	
COPD	WHR, men	324	Men	1.87 (1.62, 2.17)	4.40E-17	9.22E-06	1.84 (1.57, 2.15)	2.53E-14	6.06E-05
COPD	WHR, women	324	Women	1.24 (1.11, 1.38)	1.80E-04		1.24 (1.11, 1.39)	1.39E-04	
COPD	WHRadjBMI, men	337	Men	1.23 (1.07, 1.42)	0.004	0.007	1.11 (0.96, 1.29)	0.17	0.18
COPD	WHRadjBMI, women	337	Women	0.98 (0.89, 1.07)	0.62		0.98 (0.89, 1.08)	0.71	
Colorectal Cancer	BMI, men	565	Men	1.06 (0.91, 1.22)	0.46	0.23	1.05 (0.89, 1.25)	0.53	0.29

Outcome	GRS	N SNPs	Sex	OR (95% CI), all available cases and controls for men and women	P, all available cases and controls for men and women	P_{het}, all available cases and controls for men and women	OR (95% CI), same number of cases and controls for men and women	P, same number of cases and controls for men and women	P_{het}, same number of cases and controls for men and women
Colorectal Cancer	BMI, women	565	Women	0.92 (0.78, 1.09)	0.34		0.93 (0.79, 1.10)	0.38	
Colorectal Cancer	WHR, men	324	Men	1.31 (1.04, 1.64)	0.02	0.13	1.22 (0.94, 1.58)	0.13	0.33
Colorectal Cancer	WHR, women	324	Women	1.05 (0.87, 1.26)	0.63		1.04 (0.87, 1.25)	0.66	
Colorectal Cancer	WHRadjBMI, men	337	Men	1.28 (1.03, 1.60)	0.02	0.25	1.23 (0.96, 1.57)	0.11	0.46
Colorectal Cancer	WHRadjBMI, women	337	Women	1.10 (0.94, 1.29)	0.25		1.10 (0.94, 1.29)	0.25	
Dementia	BMI, men	565	Men	1.18 (0.84, 1.65)	0.34	0.13	0.99 (0.67, 1.45)	0.95	0.04
Dementia	BMI, women	565	Women	1.75 (1.19, 2.55)	0.004		1.74 (1.19, 2.54)	0.004	
Dementia	WHR, men	324	Men	0.58 (0.35, 0.99)	0.04	0.20	0.57 (0.32, 1.04)	0.07	0.22
Dementia	WHR, women	324	Women	0.91 (0.60, 1.39)	0.67		0.91 (0.59, 1.38)	0.64	
Dementia	WHRadjBMI, men	337	Men	0.60 (0.36, 0.98)	0.04	0.30	0.66 (0.37, 1.18)	0.16	0.53
Dementia	WHRadjBMI, women	337	Women	0.83 (0.57, 1.19)	0.31		0.82 (0.57, 1.19)	0.30	
Infertility	BMI, men	565	Men	0.68 (0.28, 1.63)	0.39	0.35	0.68 (0.28, 1.63)	0.39	0.12
Infertility	BMI, women	565	Women	1.04 (0.85, 1.28)	0.67		1.83 (0.77, 4.37)	0.17	
Infertility	WHR, men	324	Men	2.14 (0.55, 8.39)	0.27	0.30	2.14 (0.55, 8.39)	0.27	0.61
Infertility	WHR, women	324	Women	1.03 (0.83, 1.30)	0.77		1.38 (0.52, 3.64)	0.52	
Infertility	WHRadjBMI, men	337	Men	1.37 (0.37, 5.08)	0.64	0.62	1.37 (0.37, 5.08)	0.64	0.87
Infertility	WHRadjBMI, women	337	Women	0.98 (0.81, 1.19)	0.86		1.57 (0.68, 3.62)	0.29	
Lung Cancer	BMI, men	565	Men	1.19 (0.96, 1.47)	0.11	0.16	1.25 (1.00, 1.58)	0.05	0.29
Lung Cancer	BMI, women	565	Women	1.48 (1.18, 1.87)	7.01E-04		1.49 (1.19, 1.88)	6.00E-04	
Lung Cancer	WHR, men	324	Men	1.02 (0.74, 1.42)	0.89	0.47	0.99 (0.69, 1.42)	0.95	0.40
Lung Cancer	WHR, women	324	Women	1.19 (0.93, 1.54)	0.17		1.20 (0.93, 1.54)	0.17	

Outcome	GRS	N SNPs	Sex	OR (95% CI), all available cases and controls for men and women	P, all available cases and controls for men and women	P_{het}, all available cases and controls for men and women	OR (95% CI), same number of cases and controls for men and women	P, same number of cases and controls for men and women	P_{het}, same number of cases and controls for men and women
Lung Cancer	WHRadjBMI, men	337	Men	0.79 (0.57, 1.08)	0.14	0.20	0.79 (0.56, 1.12)	0.19	0.24
Lung Cancer	WHRadjBMI, women	337	Women	1.01 (0.81, 1.26)	0.92		1.01 (0.81, 1.26)	0.90	
NAFLD	BMI, men	565	Men	2.71 (2.08, 3.55)	2.43E-13	0.54	2.97 (2.22, 3.96)	1.75E-13	0.86
NAFLD	BMI, women	565	Women	3.07 (2.31, 4.10)	2.04E-14		3.08 (2.31, 4.10)	1.96E-14	
NAFLD	WHR, men	324	Men	2.40 (1.58, 3.64)	4.28E-05	0.70	2.39 (1.52, 3.75)	1.68E-04	0.71
NAFLD	WHR, women	324	Women	2.66 (1.93, 3.66)	2.06E-09		2.65 (1.93, 3.65)	2.16E-09	
NAFLD	WHRadjBMI, men	337	Men	1.31 (0.88, 1.96)	0.18	0.41	1.30 (0.84, 2.00)	0.24	0.41
NAFLD	WHRadjBMI, women	337	Women	1.61 (1.22, 2.12)	7.29E-04		1.61 (1.22, 2.12)	7.04E-04	
Renal Failure	BMI, men	565	Men	1.61 (1.44, 1.79)	1.58E-17	0.26	1.56 (1.37, 1.78)	3.07E-11	0.20
Renal Failure	BMI, women	565	Women	1.77 (1.56, 2.02)	6.08E-18		1.76 (1.54, 2.00)	2.28E-17	
Renal Failure	WHR, men	324	Men	1.94 (1.63, 2.30)	2.88E-14	5.32E-04	1.99 (1.62, 2.44)	4.63E-11	7.83E-04
Renal Failure	WHR, women	324	Women	1.31 (1.13, 1.51)	2.95E-04		1.29 (1.12, 1.50)	4.93E-04	
Renal Failure	WHRadjBMI, men	337	Men	1.31 (1.11, 1.54)	0.001	0.05	1.34 (1.11, 1.63)	0.003	0.05
Renal Failure	WHRadjBMI, women	337	Women	1.06 (0.94, 1.20)	0.34		1.06 (0.94, 1.20)	0.35	
Renal Failure - Acute	BMI, men	565	Men	1.73 (1.49, 2.01)	3.40E-13	0.32	1.75 (1.44, 2.14)	3.63E-08	0.46
Renal Failure - Acute	BMI, women	565	Women	1.97 (1.61, 2.40)	2.29E-11		1.95 (1.60, 2.38)	4.06E-11	
Renal Failure - Acute	WHR, men	324	Men	1.87 (1.48, 2.35)	1.21E-07	0.01	1.61 (1.18, 2.20)	0.003	0.18
Renal Failure - Acute	WHR, women	324	Women	1.25 (1.00, 1.56)	0.05		1.24 (0.99, 1.55)	0.06	
Renal Failure - Acute	WHRadjBMI, men	337	Men	1.18 (0.94, 1.47)	0.15	0.54	0.87 (0.65, 1.18)	0.38	0.26
Renal Failure - Acute	WHRadjBMI, women	337	Women	1.07 (0.89, 1.30)	0.47		1.07 (0.89, 1.30)	0.47	
Renal Failure - Chronic	BMI, men	565	Men	1.87 (1.59, 2.19)	2.54E-14	0.71	2.00 (1.67, 2.40)	5.25E-14	0.33

Outcome	GRS	N SNPs	Sex	OR (95% CI), all available cases and controls for men and women	P, all available cases and controls for men and women	P_{het}, all available cases and controls for men and women	OR (95% CI), same number of cases and controls for men and women	P, same number of cases and controls for men and women	P_{het}, same number of cases and controls for men and women
Renal Failure - Chronic	BMI, women	565	Women	1.78 (1.49, 2.13)	3.05E-10		1.76 (1.47, 2.11)	6.35E-10	
Renal Failure - Chronic	WHR, men	324	Men	2.33 (1.82, 3.00)	3.36E-11	1.33E-04	2.26 (1.70, 3.00)	1.65E-08	6.46E-04
Renal Failure - Chronic	WHR, women	324	Women	1.25 (1.02, 1.53)	0.03		1.24 (1.01, 1.51)	0.04	
Renal Failure - Chronic	WHRadjBMI, men	337	Men	1.67 (1.31, 2.12)	3.01E-05	3.55E-04	1.63 (1.25, 2.14)	3.81E-04	0.001
Renal Failure - Chronic	WHRadjBMI, women	337	Women	0.97 (0.82, 1.16)	0.75		0.97 (0.82, 1.15)	0.73	
Stroke	BMI, men	565	Men	1.42 (1.28, 1.57)	3.79E-11	0.84	1.40 (1.23, 1.58)	9.66E-08	0.95
Stroke	BMI, women	565	Women	1.39 (1.23, 1.57)	8.55E-08		1.40 (1.24, 1.58)	5.34E-08	
Stroke	WHR, men	324	Men	1.57 (1.33, 1.84)	5.78E-08	0.009	1.55 (1.28, 1.87)	8.55E-06	0.02
Stroke	WHR, women	324	Women	1.18 (1.03, 1.35)	0.02		1.18 (1.03, 1.35)	0.02	
Stroke	WHRadjBMI, men	337	Men	1.22 (1.05, 1.42)	0.01	0.40	1.31 (1.09, 1.58)	0.004	0.15
Stroke	WHRadjBMI, women	337	Women	1.12 (1.00, 1.26)	0.05		1.12 (0.99, 1.26)	0.06	
Stroke - Haemorrhagic	BMI, men	565	Men	1.24 (0.95, 1.62)	0.11	0.90	1.24 (0.95, 1.62)	0.11	0.86
Stroke - Haemorrhagic	BMI, women	565	Women	1.27 (0.98, 1.64)	0.07		1.29 (0.99, 1.67)	0.06	
Stroke - Haemorrhagic	WHR, men	324	Men	1.38 (0.91, 2.09)	0.13	0.54	1.38 (0.91, 2.09)	0.13	0.61
Stroke - Haemorrhagic	WHR, women	324	Women	1.18 (0.89, 1.57)	0.25		1.21 (0.90, 1.63)	0.20	
Stroke - Haemorrhagic	WHRadjBMI, men	337	Men	1.11 (0.74, 1.64)	0.62	0.99	1.11 (0.74, 1.64)	0.62	0.95
Stroke - Haemorrhagic	WHRadjBMI, women	337	Women	1.10 (0.86, 1.41)	0.44		1.12 (0.87, 1.45)	0.36	
Stroke - Ischaemic	BMI, men	565	Men	1.43 (1.20, 1.70)	6.55E-05	0.56	1.55 (1.22, 1.96)	2.92E-04	0.35
Stroke - Ischaemic	BMI, women	565	Women	1.31 (1.04, 1.65)	0.02		1.32 (1.04, 1.67)	0.02	
Stroke - Ischaemic	WHR, men	324	Men	1.33 (1.01, 1.74)	0.04	0.33	1.37 (0.95, 1.99)	0.09	0.33
Stroke - Ischaemic	WHR, women	324	Women	1.10 (0.85, 1.43)	0.47		1.10 (0.85, 1.42)	0.48	

Outcome	GRS	N SNPs	Sex	OR (95% CI), all available cases and controls for men and women	P, all available cases and controls for men and women	P_{het}, all available cases and controls for men and women	OR (95% CI), same number of cases and controls for men and women	P, same number of cases and controls for men and women	P_{het}, same number of cases and controls for men and women
Stroke - Ischaemic	WHRadjBMI, men	337	Men	1.27 (0.98, 1.65)	0.07	0.72	1.14 (0.80, 1.62)	0.46	0.85
Stroke - Ischaemic	WHRadjBMI, women	337	Women	1.19 (0.95, 1.49)	0.12		1.19 (0.95, 1.49)	0.14	
Type 1 Diabetes	BMI, men	565	Men	1.46 (1.10, 1.93)	0.009	0.23	1.39 (1.02, 1.89)	0.04	0.18
Type 1 Diabetes	BMI, women	565	Women	1.89 (1.39, 2.57)	5.45E-05		1.87 (1.37, 2.54)	6.79E-05	
Type 1 Diabetes	WHR, men	324	Men	1.15 (0.74, 1.78)	0.54	0.81	0.97 (0.60, 1.58)	0.90	0.74
Type 1 Diabetes	WHR, women	324	Women	1.07 (0.76, 1.51)	0.69		1.07 (0.76, 1.51)	0.69	
Type 1 Diabetes	WHRadjBMI, men	337	Men	1.29 (0.85, 1.97)	0.23	0.22	1.26 (0.79, 2.00)	0.33	0.29
Type 1 Diabetes	WHRadjBMI, women	337	Women	0.94 (0.70, 1.26)	0.67		0.94 (0.70, 1.26)	0.67	
Type 2 Diabetes	BMI, men	565	Men	2.73 (2.53, 2.95)	1.68E-142	3.45E-05	2.55 (2.30, 2.82)	6.42E-72	5.55E-06
Type 2 Diabetes	BMI, women	565	Women	3.58 (3.23, 3.96)	2.56E-134		3.56 (3.21, 3.94)	5.85E-132	
Type 2 Diabetes	WHR, men	324	Men	3.21 (2.85, 3.63)	8.37E-80	0.10	3.01 (2.57, 3.53)	7.11E-42	0.04
Type 2 Diabetes	WHR, women	324	Women	3.68 (3.29, 4.12)	4.10E-115		3.69 (3.29, 4.13)	5.60E-114	
Type 2 Diabetes	WHRadjBMI, men	337	Men	1.90 (1.69, 2.13)	1.65E-27	0.08	1.79 (1.53, 2.08)	8.14E-14	0.04
Type 2 Diabetes	WHRadjBMI, women	337	Women	2.17 (1.97, 2.39)	1.29E-55		2.16 (1.96, 2.38)	8.44E-55	

All effect estimates per 1 sex-stratified higher GRS. Multiple testing thresholds set to P<0.001, for 0.05/51 unique obesity trait-disease combinations in the study and P_{het}<0.001, for 0.05/48 unique male-female comparisons between the obesity trait-disease effect estimates. Table adapted (updated text and layout) from Censin, Peters, *et al.* (2019) and the corresponding pre-print (Censin, Bovijn, *et al.*, 2019).

Appendix Table 2.8. Sex-specific effects of obesity traits on leading causes of death, computed using MR-Egger and with directional pleiotropy evaluated using the MR-Egger intercept test.

Outcome	Exposure	Sex	N SNPs used (N SNPs GRS)	Intercept (95% CI)	P intercept	OR (95% CI)	P estimate	P_{het}
CAD	BMI	Combined	565 (565)	0.004 (0.001, 0.007)	0.004	1.33 (1.10, 1.59)	0.003	-
CAD	BMI	Men	565 (565)	0.004 (0.001, 0.007)	0.01	1.35 (1.12, 1.62)	0.002	0.6
CAD	BMI	Women	565 (565)	0.006 (0.002, 0.010)	0.005	1.24 (0.97, 1.58)	0.08	
CAD	WHR	Combined	324 (324)	0.008 (0.003, 0.012)	0.001	1.13 (0.85, 1.52)	0.40	-
CAD	WHR	Men	324 (324)	0.004 (-0.000, 0.008)	0.07	1.38 (1.02, 1.87)	0.04	0.94
CAD	WHR	Women	324 (324)	0.004 (-0.001, 0.008)	0.12	1.36 (1.07, 1.72)	0.01	
CAD	WHRadjBMI	Combined	337 (337)	-0.001 (-0.005, 0.003)	0.67	1.45 (1.19, 1.78)	2.99E-04	-
CAD	WHRadjBMI	Men	337 (337)	0.002 (-0.002, 0.006)	0.29	1.25 (0.97, 1.60)	0.08	0.49
CAD	WHRadjBMI	Women	337 (337)	-0.002 (-0.006, 0.002)	0.43	1.39 (1.16, 1.66)	2.72E-04	
CLD	BMI	Combined	565 (565)	0.008 (-0.002, 0.018)	0.11	1.05 (0.57, 1.92)	0.88	-
CLD	BMI	Men	565 (565)	0.009 (-0.003, 0.020)	0.13	0.87 (0.43, 1.74)	0.69	0.32
CLD	BMI	Women	565 (565)	0.005 (-0.010, 0.021)	0.51	1.56 (0.63, 3.91)	0.34	
CLD	WHR	Combined	323 (324)	-0.003 (-0.016, 0.011)	0.70	2.25 (0.96, 5.24)	0.06	-
COPD	BMI	Combined	565 (565)	0.003 (-0.000, 0.007)	0.08	1.40 (1.11, 1.77)	0.004	-
COPD	BMI	Men	565 (565)	0.008 (0.004, 0.012)	2.10E-04	1.03 (0.80, 1.32)	0.83	0.01
COPD	BMI	Women	565 (565)	0.000 (-0.005, 0.006)	0.88	1.73 (1.26, 2.36)	6.49E-04	
COPD	WHR	Combined	324 (324)	0.006 (0.001, 0.011)	0.02	1.04 (0.75, 1.44)	0.82	-
COPD	WHR	Men	324 (324)	-0.002 (-0.007, 0.002)	0.30	2.22 (1.60, 3.07)	1.77E-06	1.68E-06
COPD	WHR	Women	324 (324)	0.011 (0.005, 0.016)	1.74E-04	0.77 (0.58, 1.02)	0.07	
Lung Cancer	BMI	Combined	565 (565)	0.006 (-0.001, 0.014)	0.11	0.95 (0.60, 1.51)	0.84	-
Lung Cancer	BMI	Men	565 (565)	0.007 (-0.002, 0.016)	0.12	0.82 (0.48, 1.41)	0.47	0.46
Lung Cancer	BMI	Women	565 (565)	0.005 (-0.006, 0.016)	0.35	1.13 (0.60, 2.12)	0.71	
NAFLD	BMI	Combined	565 (565)	0.008 (-0.002, 0.017)	0.10	1.91 (1.08, 3.39)	0.03	-
NAFLD	BMI	Men	565 (565)	0.000 (-0.011, 0.011)	0.98	2.76 (1.44, 5.29)	0.002	0.08

Outcome	Exposure	Sex	N SNPs used (N SNPs GRS)	Intercept (95% CI)	P intercept	OR (95% CI)	P estimate	P_{het}
NAFLD	BMI	Women	565 (565)	0.019 (0.005, 0.033)	0.007	1.09 (0.48, 2.47)	0.83	
NAFLD	WHR	Combined	324 (324)	0.003 (-0.010, 0.016)	0.68	2.32 (1.03, 5.24)	0.04	-
NAFLD	WHR	Men	323 (324)	0.000 (-0.013, 0.013)	0.95	2.40 (0.94, 6.13)	0.07	0.94
NAFLD	WHR	Women	324 (324)	0.004 (-0.011, 0.019)	0.64	2.28 (1.06, 4.93)	0.04	
NAFLD	WHRadjBMI	Combined	337 (337)	0.002 (-0.009, 0.013)	0.74	1.51 (0.83, 2.74)	0.18	-
NAFLD	WHRadjBMI	Men	336 (337)	0.010 (-0.002, 0.022)	0.09	0.75 (0.34, 1.68)	0.49	0.26
NAFLD	WHRadjBMI	Women	337 (337)	0.005 (-0.009, 0.019)	0.47	1.34 (0.73, 2.43)	0.34	
Renal Failure	BMI	Combined	565 (565)	0.003 (-0.001, 0.007)	0.17	1.45 (1.14, 1.85)	0.003	-
Renal Failure	BMI	Men	565 (565)	0.002 (-0.002, 0.007)	0.32	1.44 (1.09, 1.90)	0.01	0.57
Renal Failure	BMI	Women	565 (565)	0.002 (-0.004, 0.007)	0.57	1.64 (1.16, 2.31)	0.005	
Renal Failure	WHR	Combined	324 (324)	0.007 (0.002, 0.013)	0.006	1.06 (0.76, 1.47)	0.73	-
Renal Failure	WHR	Men	324 (324)	0.009 (0.004, 0.014)	7.77E-04	1.14 (0.79, 1.63)	0.48	0.35
Renal Failure	WHR	Women	324 (324)	0.008 (0.002, 0.015)	0.01	0.90 (0.64, 1.27)	0.55	
Renal Failure - Acute	BMI	Combined	565 (565)	0.001 (-0.005, 0.007)	0.72	1.74 (1.23, 2.45)	0.002	-
Renal Failure - Acute	BMI	Men	565 (565)	0.001 (-0.005, 0.007)	0.78	1.67 (1.14, 2.45)	0.008	0.71
Renal Failure - Acute	BMI	Women	565 (565)	0.001 (-0.008, 0.010)	0.84	1.90 (1.12, 3.21)	0.02	
Renal Failure - Acute	WHR	Combined	324 (324)	0.008 (0.001, 0.016)	0.03	0.97 (0.61, 1.54)	0.88	-
Renal Failure - Acute	WHR	Men	324 (324)	0.006 (-0.001, 0.013)	0.09	1.29 (0.79, 2.12)	0.31	0.10
Renal Failure - Acute	WHR	Women	324 (324)	0.012 (0.003, 0.022)	0.01	0.72 (0.44, 1.18)	0.20	
Renal Failure - Chronic	BMI	Combined	565 (565)	0.005 (-0.000, 0.011)	0.05	1.36 (0.98, 1.91)	0.07	-
Renal Failure - Chronic	BMI	Men	565 (565)	0.002 (-0.005, 0.008)	0.61	1.75 (1.17, 2.61)	0.006	0.24
Renal Failure - Chronic	BMI	Women	565 (565)	0.007 (-0.001, 0.015)	0.07	1.20 (0.75, 1.94)	0.45	
Renal Failure - Chronic	WHR	Combined	324 (324)	0.008 (0.001, 0.016)	0.03	1.08 (0.68, 1.72)	0.74	-
Renal Failure - Chronic	WHR	Men	324 (324)	0.009 (0.002, 0.016)	0.02	1.36 (0.80, 2.29)	0.25	0.10
Renal Failure - Chronic	WHR	Women	324 (324)	0.011 (0.002, 0.020)	0.01	0.77 (0.49, 1.20)	0.24	

Outcome	Exposure	Sex	N SNPs used (N SNPs GRS)	Intercept (95% CI)	P intercept	OR (95% CI)	P estimate	P_{het}
Renal Failure - Chronic	WHRadjBMI	Men	337 (337)	0.002 (-0.005, 0.009)	0.58	1.50 (0.95, 2.35)	0.08	0.13
Renal Failure - Chronic	WHRadjBMI	Women	337 (337)	0.000 (-0.008, 0.009)	0.91	0.96 (0.68, 1.37)	0.83	
Stroke	BMI	Combined	565 (565)	0.005 (0.001, 0.009)	0.01	1.08 (0.86, 1.36)	0.50	-
Stroke	BMI	Men	565 (565)	0.003 (-0.002, 0.007)	0.26	1.25 (0.96, 1.64)	0.10	0.33
Stroke	BMI	Women	565 (565)	0.006 (0.000, 0.012)	0.04	1.01 (0.72, 1.42)	0.94	
Stroke	WHR	Combined	324 (324)	0.001 (-0.005, 0.006)	0.80	1.31 (0.95, 1.82)	0.10	-
Stroke	WHR	Men	324 (324)	-0.002 (-0.007, 0.003)	0.37	1.82 (1.30, 2.55)	4.75E-04	0.003
Stroke	WHR	Women	324 (324)	0.006 (-0.000, 0.012)	0.06	0.90 (0.65, 1.25)	0.54	
Stroke - Ischaemic	BMI	Combined	565 (565)	0.004 (-0.003, 0.010)	0.27	1.15 (0.78, 1.70)	0.49	-
Stroke - Ischaemic	BMI	Men	565 (565)	0.002 (-0.006, 0.009)	0.68	1.34 (0.87, 2.06)	0.19	0.40
Stroke - Ischaemic	BMI	Women	565 (565)	0.006 (-0.005, 0.016)	0.28	0.97 (0.52, 1.80)	0.91	
Type 1 Diabetes	BMI	Combined	565 (565)	-0.008 (-0.018, 0.002)	0.12	2.57 (1.41, 4.68)	0.002	-
Type 1 Diabetes	BMI	Men	565 (565)	-0.001 (-0.013, 0.010)	0.80	1.56 (0.78, 3.09)	0.21	0.09
Type 1 Diabetes	BMI	Women	565 (565)	-0.013 (-0.027, 0.001)	0.06	3.88 (1.71, 8.82)	0.001	
Type 2 Diabetes	BMI	Combined	565 (565)	-0.001 (-0.007, 0.005)	0.74	3.32 (2.35, 4.69)	1.13E-11	-
Type 2 Diabetes	BMI	Men	565 (565)	0.004 (-0.001, 0.009)	0.13	2.24 (1.64, 3.07)	4.35E-07	0.001
Type 2 Diabetes	BMI	Women	565 (565)	-0.006 (-0.013, 0.001)	0.07	5.15 (3.44, 7.73)	2.22E-15	
Type 2 Diabetes	WHR	Combined	324 (324)	0.008 (0.002, 0.015)	0.02	2.27 (1.50, 3.43)	1.09E-04	-
Type 2 Diabetes	WHR	Men	324 (324)	0.004 (-0.002, 0.010)	0.15	2.52 (1.68, 3.78)	7.89E-06	0.61
Type 2 Diabetes	WHR	Women	324 (324)	0.012 (0.005, 0.018)	4.51E-04	2.20 (1.57, 3.08)	3.87E-06	
Type 2 Diabetes	WHRadjBMI	Combined	337 (337)	0.005 (-0.002, 0.013)	0.17	1.67 (1.11, 2.49)	0.01	-
Type 2 Diabetes	WHRadjBMI	Men	337 (337)	0.008 (0.001, 0.015)	0.02	1.20 (0.76, 1.89)	0.42	0.03
Type 2 Diabetes	WHRadjBMI	Women	337 (337)	-0.001 (-0.008, 0.007)	0.85	2.24 (1.64, 3.06)	3.61E-07	

All effect estimates per 1 sex-stratified SD unit higher obesity trait. Multiple testing thresholds set to P<0.001, for 0.05/51 unique obesity trait-disease combinations in the study and P_{het}<0.001, for 0.05/48 unique male-female comparisons between the obesity trait-disease effect estimates. Table adapted (updated text and layout) from Censin, Peters, *et al.* (2019) and the corresponding pre-print (Censin, Bovijn, *et al.*, 2019).

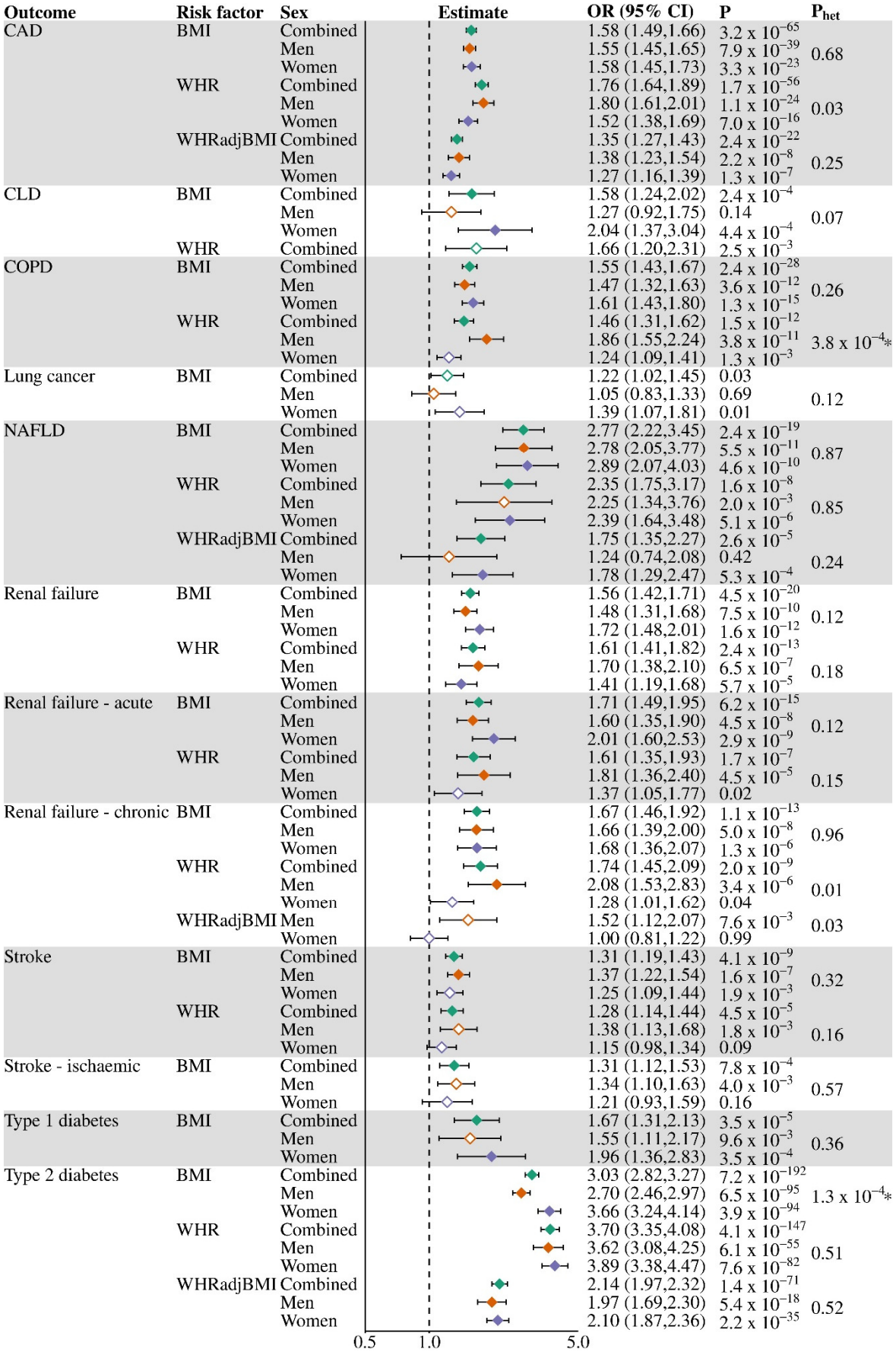
Appendix Table 2.9. Sex-specific effects of obesity traits on fasting glucose and fasting insulin, computed using MR.

Exposure	Outcome	Sex	N SNPs available (N SNPs GRS)	Outcome Unit	Method	Estimate (95% CI)	P	P_{het}
BMI	FG	Combined	482 (565)	mmol/L	IVW	0.075 (0.054,0.095)	6.72E-13	-
BMI	FG	Combined	482 (565)	mmol/L	MR-Egger	0.102 (0.049,0.156)	1.85E-04	-
BMI	FG	Combined	482 (565)	mmol/L	MR-Egger (intercept)	-0.000 (-0.001,0.000)	0.27	-
BMI	FG	Combined	482 (565)	mmol/L	Weighted median	0.078 (0.050,0.107)	7.71E-08	-
BMI	FG	Men	482 (565)	mmol/L	IVW	0.066 (0.039,0.092)	1.51E-06	0.42
BMI	FG	Men	482 (565)	mmol/L	MR-Egger	0.060 (-0.003,0.124)	0.06	0.11
BMI	FG	Men	482 (565)	mmol/L	MR-Egger (intercept)	0.000 (-0.001,0.001)	0.86	0.16
BMI	FG	Men	482 (565)	mmol/L	Weighted median	0.074 (0.025,0.123)	0.003	0.26
BMI	FG	Women	482 (565)	mmol/L	IVW	0.081 (0.056,0.106)	3.51E-10	0.42
BMI	FG	Women	482 (565)	mmol/L	MR-Egger	0.136 (0.069,0.202)	5.95E-05	0.11
BMI	FG	Women	482 (565)	mmol/L	MR-Egger (intercept)	-0.001 (-0.002,0.000)	0.08	0.16
BMI	FG	Women	482 (565)	mmol/L	Weighted median	0.109 (0.073,0.145)	3.20E-09	0.26
BMI	FI	Combined	482 (565)	ln(pmol/L)	IVW	0.214 (0.191,0.236)	2.25E-78	-
BMI	FI	Combined	482 (565)	ln(pmol/L)	MR-Egger	0.214 (0.155,0.273)	1.17E-12	-
BMI	FI	Combined	482 (565)	ln(pmol/L)	MR-Egger (intercept)	-0.000 (-0.001,0.001)	0.99	-
BMI	FI	Combined	482 (565)	ln(pmol/L)	Weighted median	0.222 (0.184,0.260)	7.89E-31	-
BMI	FI	Men	482 (565)	ln(pmol/L)	IVW	0.222 (0.191,0.253)	2.94E-44	0.50
BMI	FI	Men	482 (565)	ln(pmol/L)	MR-Egger	0.253 (0.179,0.326)	1.67E-11	0.21
BMI	FI	Men	482 (565)	ln(pmol/L)	MR-Egger (intercept)	-0.001 (-0.002,0.001)	0.36	0.29
BMI	FI	Men	482 (565)	ln(pmol/L)	Weighted median	0.221 (0.166,0.276)	2.52E-15	0.89
BMI	FI	Women	482 (565)	ln(pmol/L)	IVW	0.207 (0.178,0.236)	2.36E-44	0.50
BMI	FI	Women	482 (565)	ln(pmol/L)	MR-Egger	0.185 (0.109,0.261)	1.96E-06	0.21
BMI	FI	Women	482 (565)	ln(pmol/L)	MR-Egger (intercept)	0.000 (-0.001,0.002)	0.55	0.29
BMI	FI	Women	482 (565)	ln(pmol/L)	Weighted median	0.216 (0.167,0.264)	2.77E-18	0.89

Exposure	Outcome	Sex	N SNPs available (N SNPs GRS)	Outcome Unit	Method	Estimate (95% CI)	P	P_{het}
WHR	FG	Combined	261 (324)	mmol/L	IVW	0.078 (0.050,0.106)	4.54E-08	-
WHR	FG	Combined	261 (324)	mmol/L	MR-Egger	0.077 (-0.015,0.169)	0.10	-
WHR	FG	Combined	261 (324)	mmol/L	MR-Egger (intercept)	0.000 (-0.001,0.001)	0.98	-
WHR	FG	Combined	261 (324)	mmol/L	Weighted median	0.075 (0.038,0.113)	7.26E-05	-
WHR	FG	Men	261 (324)	mmol/L	IVW	0.061 (0.019,0.102)	0.004	0.41
WHR	FG	Men	261 (324)	mmol/L	MR-Egger	0.083 (-0.006,0.172)	0.07	0.84
WHR	FG	Men	261 (324)	mmol/L	MR-Egger (intercept)	-0.000 (-0.002,0.001)	0.58	0.93
WHR	FG	Men	261 (324)	mmol/L	Weighted median	0.080 (0.015,0.145)	0.02	0.88
WHR	FG	Women	261 (324)	mmol/L	IVW	0.082 (0.053,0.111)	4.21E-08	0.41
WHR	FG	Women	261 (324)	mmol/L	MR-Egger	0.095 (0.023,0.167)	0.01	0.84
WHR	FG	Women	261 (324)	mmol/L	MR-Egger (intercept)	-0.000 (-0.002,0.001)	0.70	0.93
WHR	FG	Women	261 (324)	mmol/L	Weighted median	0.086 (0.044,0.128)	5.77E-05	0.88
WHR	FI	Combined	261 (324)	ln(pmol/L)	IVW	0.217 (0.179,0.254)	3.46E-29	-
WHR	FI	Combined	261 (324)	ln(pmol/L)	MR-Egger	0.212 (0.089,0.336)	7.53E-04	-
WHR	FI	Combined	261 (324)	ln(pmol/L)	MR-Egger (intercept)	0.000 (-0.002,0.002)	0.94	-
WHR	FI	Combined	261 (324)	ln(pmol/L)	Weighted median	0.223 (0.176,0.271)	3.17E-20	-
WHR	FI	Men	261 (324)	ln(pmol/L)	IVW	0.195 (0.137,0.253)	3.65E-11	0.82
WHR	FI	Men	261 (324)	ln(pmol/L)	MR-Egger	0.200 (0.075,0.324)	0.002	0.90
WHR	FI	Men	261 (324)	ln(pmol/L)	MR-Egger (intercept)	-0.000 (-0.002,0.002)	0.93	0.96
WHR	FI	Men	261 (324)	ln(pmol/L)	Weighted median	0.207 (0.130,0.285)	1.56E-07	0.95
WHR	FI	Women	261 (324)	ln(pmol/L)	IVW	0.203 (0.165,0.241)	4.32E-26	0.82
WHR	FI	Women	261 (324)	ln(pmol/L)	MR-Egger	0.210 (0.117,0.302)	8.76E-06	0.90
WHR	FI	Women	261 (324)	ln(pmol/L)	MR-Egger (intercept)	-0.000 (-0.002,0.002)	0.88	0.96
WHR	FI	Women	261 (324)	ln(pmol/L)	Weighted median	0.205 (0.151,0.258)	6.51E-14	0.95
WHRadjBMI	FG	Combined	267 (337)	mmol/L	IVW	0.046 (0.019,0.073)	8.38E-04	-

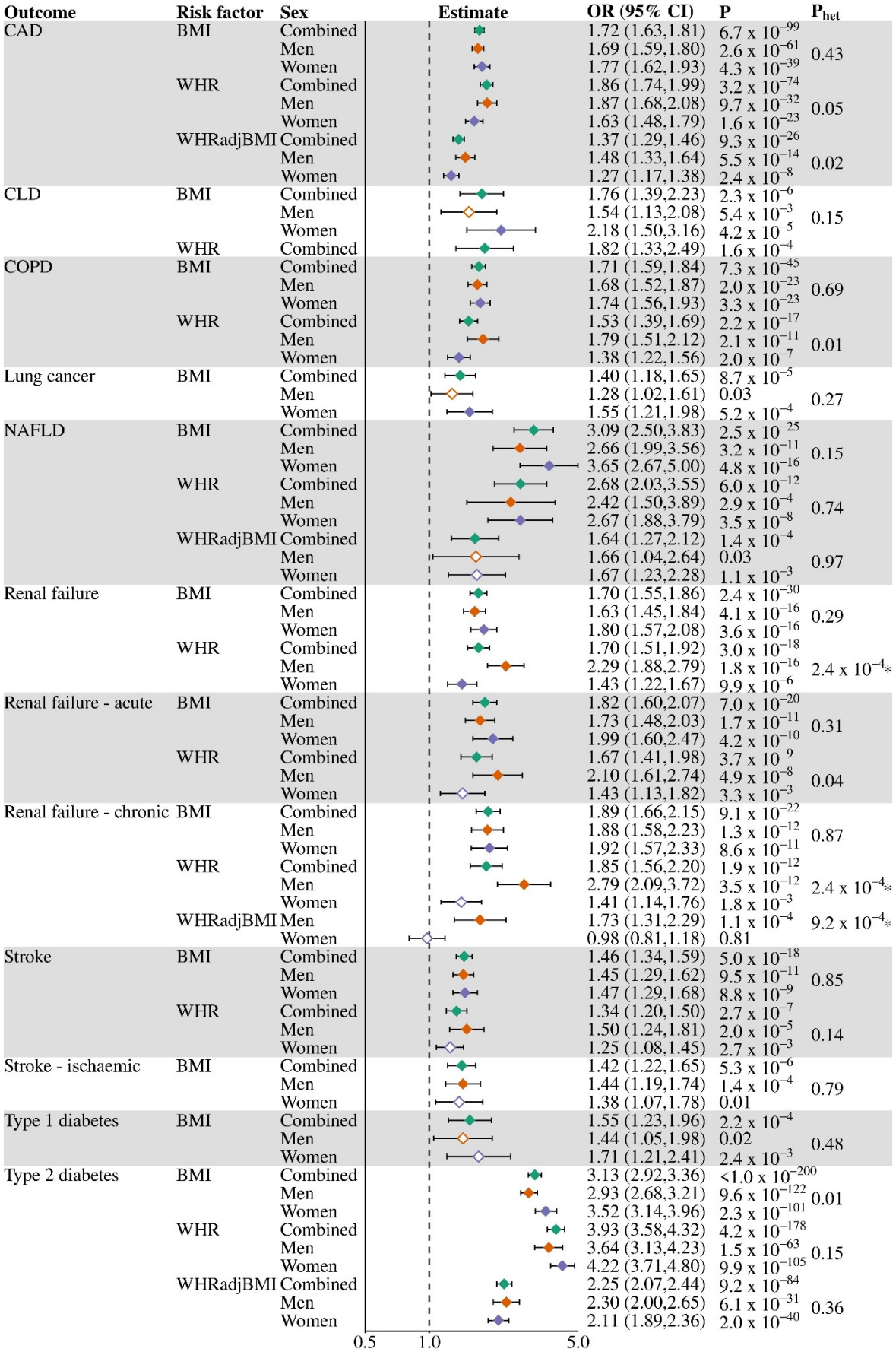
Exposure	Outcome	Sex	N SNPs available (N SNPs GRS)	Outcome Unit	Method	Estimate (95% CI)	P	P_{het}
WHRadjBMI	FG	Combined	267 (337)	mmol/L	MR-Egger	0.053 (-0.022,0.128)	0.17	-
WHRadjBMI	FG	Combined	267 (337)	mmol/L	MR-Egger (intercept)	-0.000 (-0.001,0.001)	0.84	-
WHRadjBMI	FG	Combined	267 (337)	mmol/L	Weighted median	0.040 (0.007,0.074)	0.02	-
WHRadjBMI	FG	Men	267 (337)	mmol/L	IVW	0.028 (-0.017,0.074)	0.22	0.46
WHRadjBMI	FG	Men	267 (337)	mmol/L	MR-Egger	-0.008 (-0.102,0.086)	0.87	0.06
WHRadjBMI	FG	Men	267 (337)	mmol/L	MR-Egger (intercept)	0.001 (-0.001,0.002)	0.39	0.06
WHRadjBMI	FG	Men	267 (337)	mmol/L	Weighted median	0.024 (-0.036,0.084)	0.43	0.45
WHRadjBMI	FG	Women	267 (337)	mmol/L	IVW	0.048 (0.023,0.072)	1.50E-04	0.46
WHRadjBMI	FG	Women	267 (337)	mmol/L	MR-Egger	0.096 (0.040,0.153)	7.98E-04	0.06
WHRadjBMI	FG	Women	267 (337)	mmol/L	MR-Egger (intercept)	-0.001 (-0.002,0.000)	0.06	0.06
WHRadjBMI	FG	Women	267 (337)	mmol/L	Weighted median	0.052 (0.015,0.088)	0.006	0.45
WHRadjBMI	FI	Combined	267 (337)	ln(pmol/L)	IVW	0.137 (0.104,0.170)	2.23E-16	-
WHRadjBMI	FI	Combined	267 (337)	ln(pmol/L)	MR-Egger	0.187 (0.096,0.278)	5.91E-05	-
WHRadjBMI	FI	Combined	267 (337)	ln(pmol/L)	MR-Egger (intercept)	-0.001 (-0.002,0.001)	0.26	-
WHRadjBMI	FI	Combined	267 (337)	ln(pmol/L)	Weighted median	0.141 (0.099,0.184)	7.98E-11	-
WHRadjBMI	FI	Men	267 (337)	ln(pmol/L)	IVW	0.093 (0.041,0.144)	4.32E-04	0.13
WHRadjBMI	FI	Men	267 (337)	ln(pmol/L)	MR-Egger	0.037 (-0.069,0.143)	0.49	0.005
WHRadjBMI	FI	Men	267 (337)	ln(pmol/L)	MR-Egger (intercept)	0.001 (-0.001,0.002)	0.24	0.01
WHRadjBMI	FI	Men	267 (337)	ln(pmol/L)	Weighted median	0.078 (0.006,0.150)	0.03	0.08
WHRadjBMI	FI	Women	267 (337)	ln(pmol/L)	IVW	0.140 (0.108,0.173)	2.57E-17	0.13
WHRadjBMI	FI	Women	267 (337)	ln(pmol/L)	MR-Egger	0.223 (0.149,0.296)	3.22E-09	0.005
WHRadjBMI	FI	Women	267 (337)	ln(pmol/L)	MR-Egger (intercept)	-0.002 (-0.004,-0.000)	0.02	0.01
WHRadjBMI	FI	Women	267 (337)	ln(pmol/L)	Weighted median	0.155 (0.107,0.203)	2.42E-10	0.08

All effect estimates per 1 sex-stratified SD unit higher obesity trait. Multiple testing thresholds set to P<0.003 and P_{het}<0.003, for 0.05/15 unique obesity trait-risk factors evaluated and male-female effect estimate comparisons. Table adapted (updated text and layout) from Censin, Peters, et al. (2019) and the corresponding pre-print (Censin, Bovijn, et al., 2019).



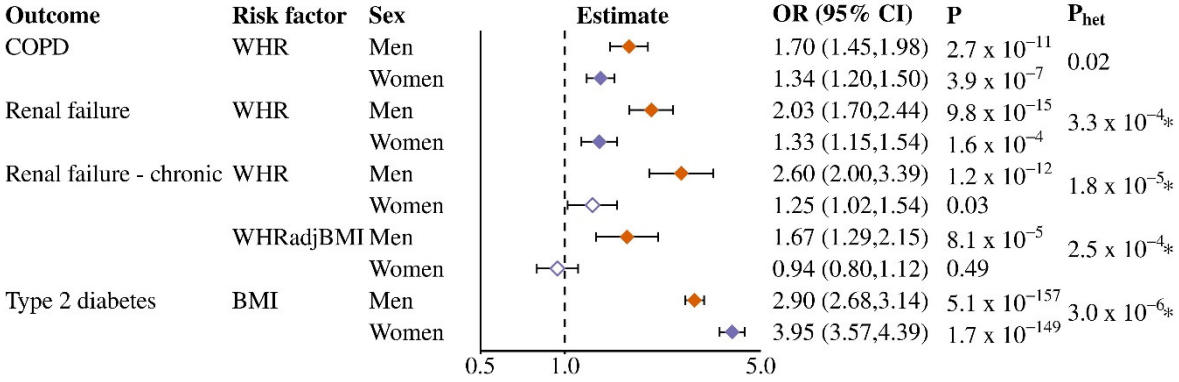
Appendix Figure 2.1. Sex-specific effects of obesity on leading causes of death using

externally weighted GRSs and the ratio method. Obesity trait-disease associations surpassing the P-value threshold (<0.001, for 0.05/51 obesity trait-disease combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.001, for 0.05/48 obesity trait-disease combinations where sexual heterogeneity in the effect estimates were compared; no test for sexual heterogeneity performed for breast cancer) are marked with *. All effect estimates per 1 sex-stratified SD unit higher obesity trait. Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Figure adapted (updated layout) from Censin, Peters, *et al.* (2019) and the corresponding pre-print (Censin, Bovijn, *et al.*, 2019).

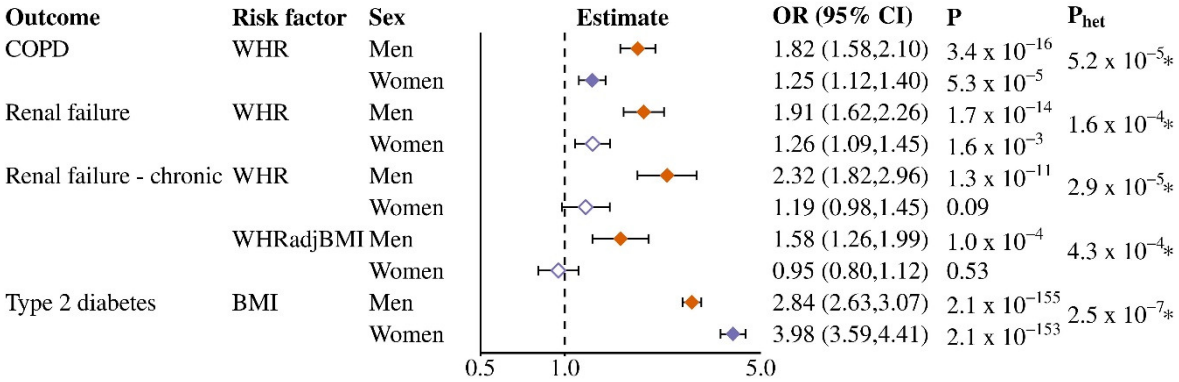


Appendix Figure 2.2. Sex-specific effects of obesity on leading causes of death using

unweighted GRSs and the ratio method. Obesity trait-disease associations surpassing the P-value threshold (<0.001, for 0.05/51 obesity trait-disease combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.001, for 0.05/48 obesity trait-disease combinations where sexual heterogeneity in the effect estimates were compared; no test for sexual heterogeneity performed for breast cancer) are marked with *. All effect estimates per 1 sex-stratified SD unit higher obesity trait. Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Figure adapted (updated layout) from Censin, Peters, *et al.* (2019) and the corresponding pre-print (Censin, Bovijn, *et al.*, 2019).

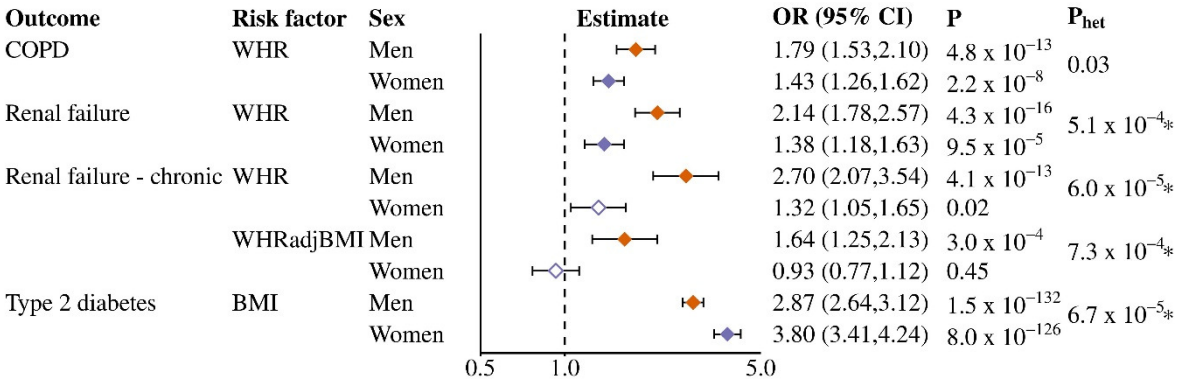


Appendix Figure 2.3. Obesity trait-disease combinations with evidence of a sex difference in the main analysis, recomputed using the Sex-combined primary SNPs and combined weights approach and with effect estimates computed using the ratio method. Obesity trait-disease associations surpassing the P-value threshold (<0.001, for 0.05/51 obesity trait-disease combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.001, for 0.05/48 obesity trait-disease combinations where sexual heterogeneity in the effect estimates were compared; no test for sexual heterogeneity performed for breast cancer) are marked with *. All effect estimates per 1 sex-stratified SD unit higher obesity trait. Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Figure adapted (updated layout) from Censin *et al.* (2019).

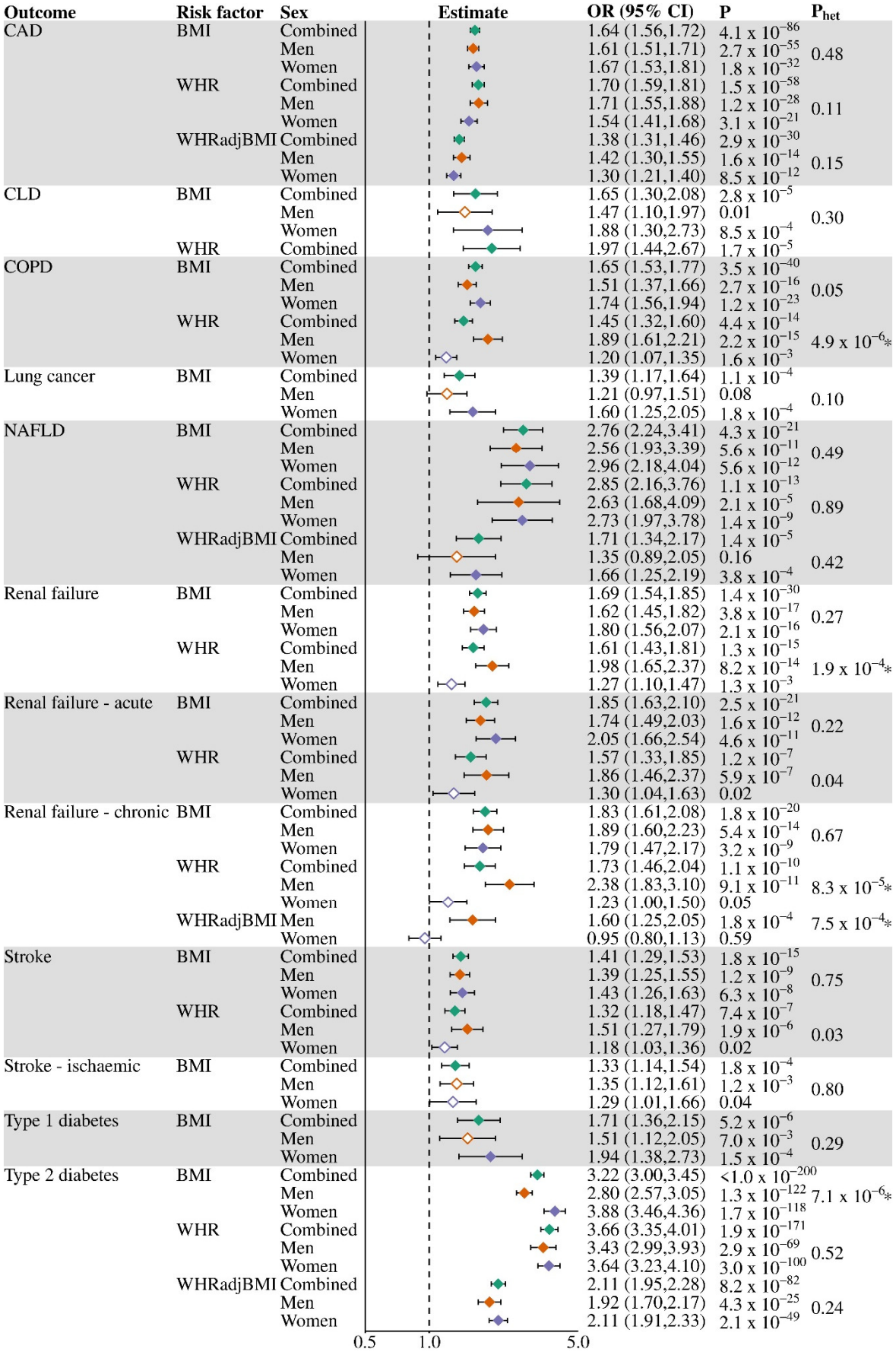


Appendix Figure 2.4. Obesity trait-disease combinations with evidence of a sex difference in the main analysis, recomputed using the Sex-combined primary SNPs and sex-specific weights approach and with effect estimates computed using the ratio method. Obesity trait-disease associations surpassing the P-value threshold (<0.001, for 0.05/51 obesity trait-disease combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.001, for 0.05/48 obesity trait-disease combinations where sexual heterogeneity in the effect estimates were compared; no test for sexual heterogeneity performed for breast cancer) are marked with *. All effect estimates per 1 sex-stratified SD unit higher obesity trait. Sex-combined effect

estimates in green, men effect estimates in orange, and women effect estimates in purple. Figure adapted (updated layout) from Censin *et al.* (2019).

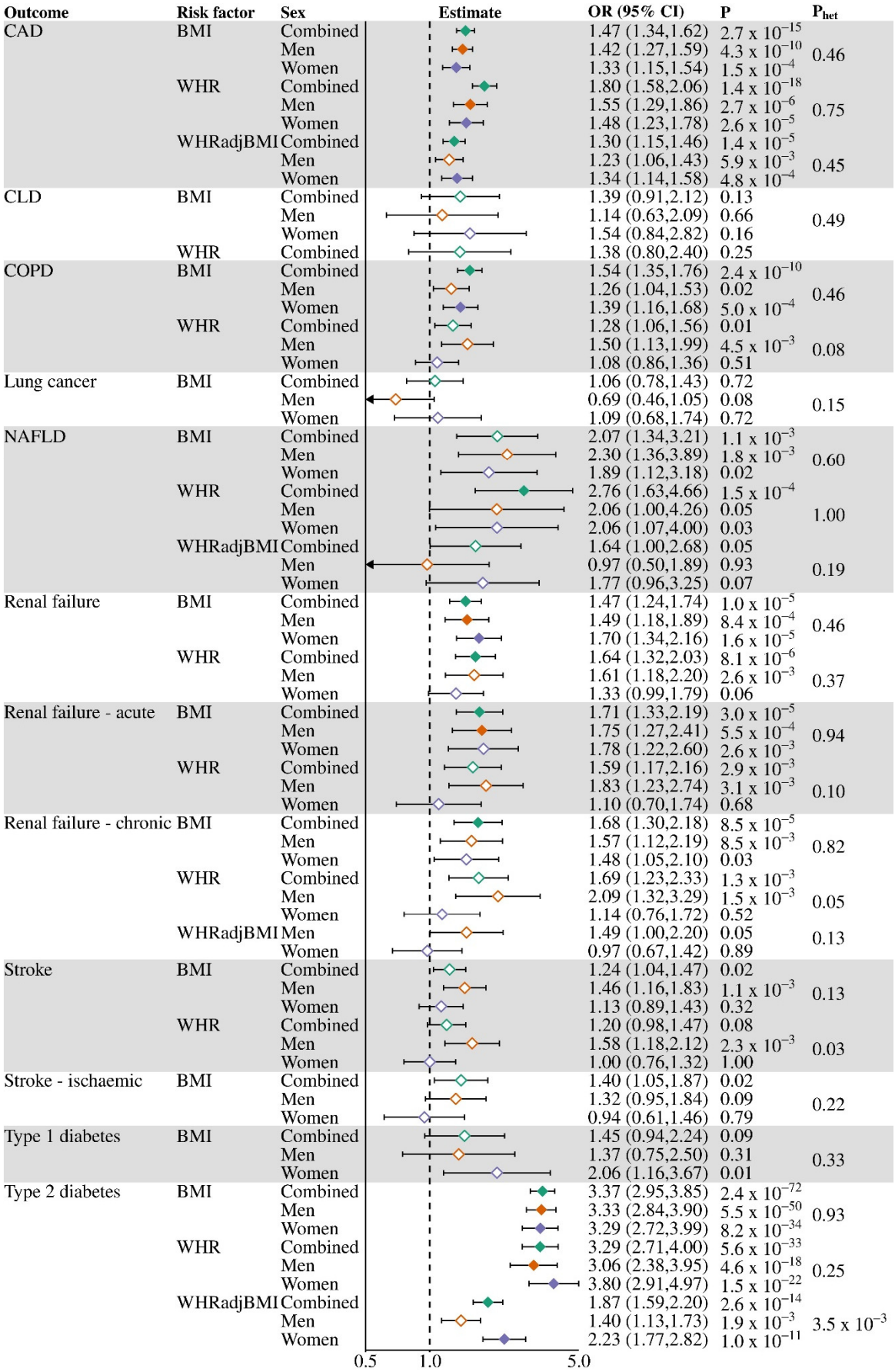


Appendix Figure 2.5. Obesity trait-disease combinations with evidence of a sex difference in the main analysis, recomputed using the Sex-combined primary SNPs and unweighted approach and with effect estimates computed using the ratio method. Obesity trait-disease associations surpassing the P-value threshold (<0.001, for 0.05/51 obesity trait-disease combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.001, for 0.05/48 obesity trait-disease combinations where sexual heterogeneity in the effect estimates were compared; no test for sexual heterogeneity performed for breast cancer) are marked with *. All effect estimates per 1 sex-stratified SD unit higher obesity trait. Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Figure adapted (updated layout) from Censin *et al.* (2019).



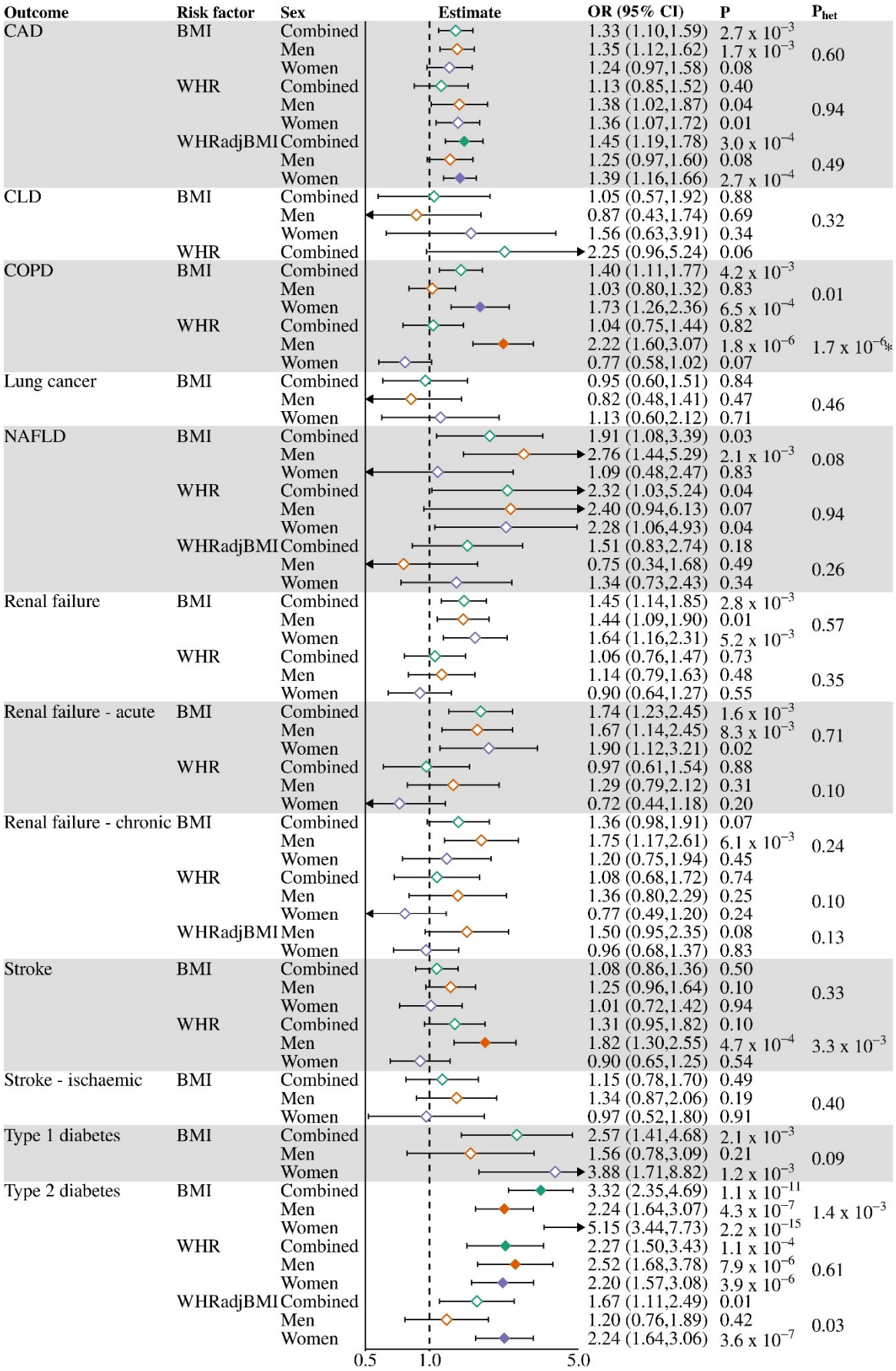
Appendix Figure 2.6. Sex-specific effects of obesity on leading causes of death in British

ancestry subset and with effect estimates computed using the ratio method. Obesity trait-disease associations surpassing the P-value threshold (<0.001, for 0.05/51 obesity trait-disease combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.001, for 0.05/48 obesity trait-disease combinations where sexual heterogeneity in the effect estimates were compared; no test for sexual heterogeneity performed for breast cancer) are marked with *. All effect estimates per 1 sex-stratified SD unit higher obesity trait. Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Figure adapted (updated layout) from Censin, Peters, *et al.* (2019) and the corresponding pre-print (Censin, Bovijn, *et al.*, 2019).



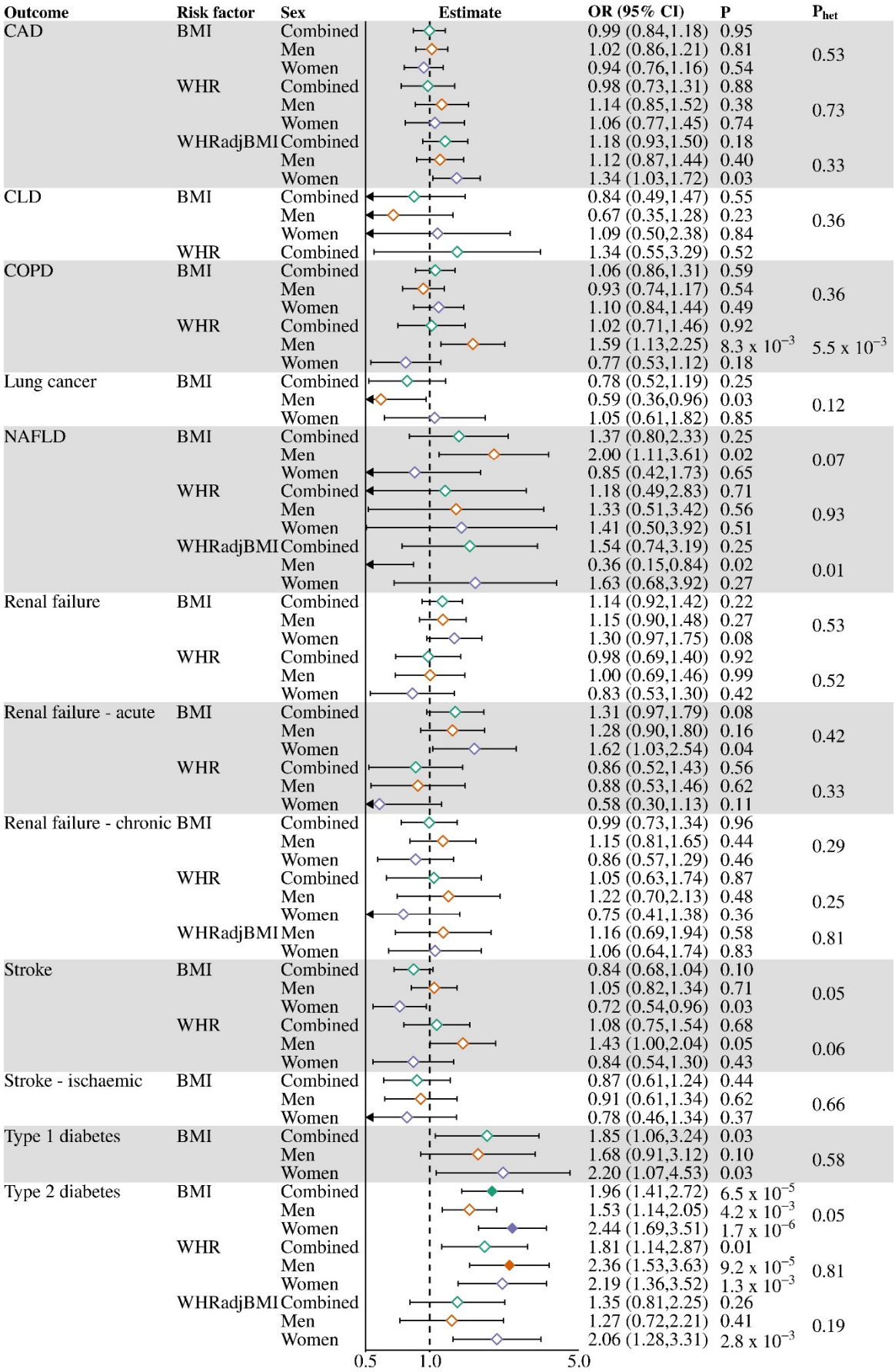
Appendix Figure 2.7. Sex-specific effects of obesity on leading causes of death using the

externally weighted approach approach and the weighted median method. Obesity trait-disease associations surpassing the P-value threshold (<0.001, for 0.05/51 obesity trait-disease combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.001, for 0.05/48 obesity trait-disease combinations where sexual heterogeneity in the effect estimates were compared; no test for sexual heterogeneity performed for breast cancer) are marked with *. All effect estimates per 1 sex-stratified SD unit higher obesity trait. Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Layout copied from figures in Censin, Peters, *et al.* (2019) and the corresponding pre-print (Censin, Bovijn, *et al.*, 2019).



Appendix Figure 2.8. Sex-specific effects of obesity on leading causes of death using the main

sex-specific effect estimates approach and the MR-Egger method. Obesity trait-disease associations surpassing the P-value threshold (<0.001, for 0.05/51 obesity trait-disease combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.001, for 0.05/48 obesity trait-disease combinations where sexual heterogeneity in the effect estimates were compared; no test for sexual heterogeneity performed for breast cancer) are marked with *. All effect estimates per 1 sex-stratified SD unit higher obesity trait. Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Layout copied from figures in Censin, Peters, *et al.* (2019) and the corresponding pre-print (Censin, Bovijn, *et al.*, 2019).



Appendix Figure 2.9. Sex-specific effects of obesity on leading causes of death using the

externally weighted approach approach and the MR-Egger method. Obesity trait-disease associations surpassing the P-value threshold (<0.001, for 0.05/51 obesity trait-disease combinations in the study) have filled diamonds, and P_{het}-values that surpass the P_{het}-threshold (<0.001, for 0.05/48 obesity trait-disease combinations where sexual heterogeneity in the effect estimates were compared; no test for sexual heterogeneity performed for breast cancer) are marked with *. All effect estimates per 1 sex-stratified SD unit higher obesity trait. Sex-combined effect estimates in green, men effect estimates in orange, and women effect estimates in purple. Layout copied from figures in Censin *et al.* (2019).

Appendix 3. Supplement to Chapter 3

Appendix Table 3.1. Basic information and sensitivity analysis statistics for the sex-differential pQTLs.

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
ACPL2.6079.59.3	2-phosphoxylose phosphatase 1	Q8TE99	rs7190621	16	85289100	T	C	0.28	0.28	trans	-	-
ADAMTS3.8845.2.3	A disintegrin and metalloproteinase with thrombospondin motifs 3	O15072	rs7277195	21	29949276	A	G	0.25	0.23	trans	-	-
ADAMTS4.2809.25.2	A disintegrin and metalloproteinase with thrombospondin motifs 4	O75173	rs200641322	7	62925296	T	G	0.94	0.94	trans	-	-
ADAMTS5.3168.8.2	A disintegrin and metalloproteinase with thrombospondin motifs 5	Q9UNA0	rs4958481	5	151029529	G	A	0.91	0.92	trans	-	-
ADCK4.9794.17.3	AarF domain-containing protein kinase 4	Q96D53	rs34962524	1	223026079	G	A	0.83	0.83	trans	-	-
ADCYAP1.4593.11.1	Pituitary adenylate cyclase-activating polypeptide 27	P18509	rs7926118	11	78438276	G	A	0.74	0.74	trans	-	-
ADCYAP1.4593.11.1	Pituitary adenylate cyclase-activating polypeptide 27	P18509	rs11100189	4	140554013	G	A	0.67	0.66	trans	-	-
ADCYAP1.8285.64.3	Pituitary adenylate cyclase-activating polypeptide	P18509	rs76713390	16	5096757	A	C	0.88	0.88	trans	-	-
ADCYAP1.8285.64.3	Pituitary adenylate cyclase-activating polypeptide	P18509	rs61660593	8	126483900	G	A	0.96	0.96	trans	-	-
ADGRF1.11243.90.3	Adhesion G-protein coupled receptor F1	Q5T601	rs6884348	5	3394590	C	T	0.72	0.72	trans	-	-
ADGRF1.11243.90.3	Adhesion G-protein coupled receptor F1	Q5T601	rs55885888	7	3800227	G	C	0.96	0.96	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
ADGRF5.6409.57.3	Adhesion G protein-coupled receptor F5	Q8IZF2	rs10053843	5	125708858	A	C	0.50	0.49	trans	-	-
ADH1B.9834.62.3	Alcohol dehydrogenase 1B	P00325	rs7808576	7	11706221	G	T	0.50	0.52	trans	-	-
ADH4.8325.37.3	Alcohol dehydrogenase 4	P08319	rs77672249	20	18889944	G	A	0.93	0.94	trans	-	-
AGER.4125.52.2	Advanced glycosylation end product-specific receptor, soluble	Q15109	rs79768452	6	153661753	A	G	0.94	0.94	trans	-	-
AKR1B1.9854.36.3	Aldose reductase	P15121	rs11640887	16	86051882	C	T	0.71	0.70	trans	-	-
ALDH3A1.11480.1.3	Aldehyde dehydrogenase, dimeric NADP-preferring	P30838	rs74418288	7	63566842	T	G	0.97	0.97	trans	-	-
ALDH3B1.12940.35.3	Aldehyde dehydrogenase family 3 member B1	P43353	rs56292770	21	46810010	G	A	0.71	0.72	trans	-	-
ANAPC7.11690.47.3	Anaphase-promoting complex subunit 7	Q9UJX3	rs3760003	16	77222505	C	T	0.42	0.42	trans	-	-
ANGPT2.13660.76.3	Angiopoietin-2	O15123	rs73127904	1	245564582	G	A	0.95	0.94	trans	-	-
ANGPT4.2500.2.3	Angiopoietin-4	Q9Y264	rs35185467	4	178687656	C	T	0.75	0.74	trans	-	-
ANGPTL1.9092.33.3	Angiopoietin-related protein 1	O95841	rs55759365	2	47114583	A	C	0.72	0.72	trans	no	-
ANKRD27.12445.50.3	Ankyrin repeat domain-containing protein 27	Q96NW4	rs535816341	12	91995501	T	A	0.97	0.97	trans	-	-
ANKRD27.12445.50.3	Ankyrin repeat domain-containing protein 27	Q96NW4	rs12547239	8	40993285	G	A	0.30	0.32	trans	-	-
ANKRD46.7851.30.3	Ankyrin repeat domain-containing protein 46	Q86W74	rs75920135	2	11051446	G	A	0.82	0.83	trans	-	-
ANP32B.4194.26.3	Acidic leucine-rich nuclear phosphoprotein 32 family member B	Q92688	rs1651667	16	57056609	A	G	0.29	0.29	trans	-	-
AP4M1.10076.1.3	AP-4 complex subunit mu-1	O00189	rs76553855	18	9558712	T	A	0.76	0.76	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
APBB2.12753.6.3	Amyloid beta A4 precursor protein-binding family B member 2	Q92870	rs80029890	7	54449520	G	T	0.96	0.96	trans	no	-
APCDD1.6599.5.3	Protein APCDD1	Q8J025	rs796500	3	111727739	A	G	0.93	0.92	trans	-	-
APCS.2474.54.5	Serum amyloid P-component	P02743	rs34827770	7	108176137	A	G	0.90	0.90	trans	-	-
APLP2.10627.87.3	Amyloid-like protein 2	Q06481	rs116928246	12	125025742	G	A	0.54	0.55	trans	-	no
APOL1.11510.31.3	Apolipoprotein L1	O14791	rs12193015	6	4620443	C	T	0.60	0.60	trans	no	-
APOM.10445.20.3	Apolipoprotein M	O95445	rs17713196	3	165489724	C	T	0.81	0.79	trans	no	-
APOM.14125.5.3	Apolipoprotein M	O95445	rs17581439	2	172737267	T	C	0.74	0.73	trans	no	-
ARFIP1.13488.3.3	Arfaptin-1	P53367	rs73047026	3	22167075	A	G	0.74	0.73	trans	-	-
ARHGAP36.6289.78.3	Rho GTPase-activating protein 36	Q6ZRI8	rs12153655	5	128207645	C	T	0.78	0.78	trans	-	-
ARHGAP36.6289.78.3	Rho GTPase-activating protein 36	Q6ZRI8	rs12339857	9	110525413	A	G	0.92	0.92	trans	-	-
ARHGEF1.13976.9.3	Rho guanine nucleotide exchange factor 1	Q92888	rs9364470	6	169005228	T	C	0.81	0.81	trans	-	-
ARID3A.3875.62.1	AT-rich interactive domain-containing protein 3A	Q99856	rs9537337	13	56775627	G	A	0.07	0.06	trans	-	-
ARID3A.3875.62.1	AT-rich interactive domain-containing protein 3A	Q99856	rs1406742	2	66324534	T	C	0.19	0.20	trans	-	-
ARRDC3.12352.70.3	Arrestin domain-containing protein 3	Q96B67	rs13018449	2	234486645	C	A	0.23	0.22	trans	-	-
ASGR1.5452.71.3	Asialoglycoprotein receptor 1	P07306	rs6431973	2	9185559	C	T	0.34	0.33	trans	-	-
ASMTL.10089.7.3	N-acetylserotonin O-methyltransferase-like protein	O95671	rs73179258	13	46485297	G	A	0.97	0.96	trans	-	-
ATAD1.6625.31.3	ATPase family AAA domain-containing protein 1	Q8NBU5	rs61923795	12	61221601	G	A	0.92	0.92	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
ATAD1.6625.31.3	ATPase family AAA domain-containing protein 1	Q8NBU5	rs474122	18	3603112	A	G	0.26	0.26	trans	-	-
ATP4B.9994.217.3	Potassium-transporting ATPase subunit beta	P51164	rs117811535	13	64038636	C	G	0.97	0.97	trans	-	-
ATP4B.9994.217.3	Potassium-transporting ATPase subunit beta	P51164	rs6502557	17	16960978	G	A	0.86	0.85	trans	-	-
AURKA.3091.70.2	Aurora kinase A	O14965	rs79714874	3	17029327	A	T	0.96	0.96	trans	-	-
AXIN2.12925.105.3	Axin-2	Q9Y2T1	rs6559981	9	89965663	T	C	0.25	0.27	trans	no	-
AXIN2.8429.16.3	Axin-2	Q9Y2T1	rs375029398	9	81428904	A	G	0.95	0.95	trans	no	-
AZGP1.9312.8.3	Zinc-alpha-2-glycoprotein	P25311	rs10077981	5	37901492	A	G	0.51	0.50	trans	-	-
B3GNT8.9297.12.3	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 8	Q7Z7M8	rs4895754	6	148759707	G	T	0.13	0.14	trans	-	-
B4GALT6.10832.24.3	Beta-1,4-galactosyltransferase 6	Q9UBX8	rs34932812	2	84996343	G	A	0.98	0.98	trans	-	-
BACH1.12451.62.3	Transcription regulator protein BACH1	O14867	rs12642529	4	38168937	G	A	0.19	0.21	trans	-	-
BAG3.10078.5.3	BAG family molecular chaperone regulator 3	O95817	rs140735987	1	76298756	A	G	0.94	0.94	trans	-	-
BCAN.3461.58.1	Brevican core protein	Q96GW7	rs13023465	2	154035508	C	T	0.82	0.81	trans	-	-
BCAN.3461.58.1	Brevican core protein	Q96GW7	rs1326906	9	78473903	T	G	0.98	0.98	trans	-	-
BCL2.3412.7.1	Apoptosis regulator Bcl-2	P10415	rs1009850	9	131193665	A	C	0.46	0.45	trans	-	-
BCL2A1.3413.50.2	Bcl-2-related protein A1	Q16548	rs13165690	5	3420778	G	A	0.70	0.70	trans	-	-
BCL6.13640.5.3	B-cell lymphoma 6 protein	P41182	rs9717986	6	162178824	C	T	0.10	0.10	trans	-	-
BET1L.10959.125.3	BET1-like protein	Q9NYM9	rs80011000	21	21436014	A	T	0.96	0.96	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
BGLAP.11067.13.3	Osteocalcin	P02818	rs17187009	1	208235103	A	G	0.94	0.94	trans	-	-
BGN.13690.26.3	Biglycan	P21810	rs13065384	3	22924055	C	T	0.82	0.81	trans	no	-
BMP10.3587.53.2	Bone morphogenetic protein 10	O95393	rs34998662	2	47327405	G	C	0.96	0.96	trans	-	-
BMP6.3736.60.3	Bone morphogenetic protein 6	P22004	rs6067937	20	50403808	C	T	0.96	0.96	trans	-	-
BMPR1A.4859.6.1	Bone morphogenetic protein receptor type-1A	P36894	rs11627959	14	35160471	A	C	0.44	0.43	trans	-	-
BNIP3L.7835.2.3	BCL2/adenovirus E1B 19 kDa protein-interacting protein 3-like	O60238	rs6840503	4	147467353	A	G	0.17	0.16	trans	-	-
BRSK2.9790.28.3	Serine/threonine-protein kinase BRSK2	Q8IWQ3	rs661325	6	167181032	C	G	0.37	0.38	trans	-	-
BTNL9.7950.142.3	Butyrophilin-like protein 9	Q6UXG8	rs71385947	17	76321675	G	A	0.89	0.91	trans	-	-
C17orf78.6571.75.3	Uncharacterised protein C17orf78	Q8N4C9	rs34164740	1	179295760	T	C	0.76	0.76	trans	-	no
C17orf78.6571.75.3	Uncharacterised protein C17orf78	Q8N4C9	rs6074977	20	15931275	C	A	0.03	0.03	trans	no	-
C18orf32.8236.8.3	UPF0729 protein C18orf32	Q8TCD1	rs114441192	1	93090853	C	T	0.96	0.97	trans	-	-
C1orf198.8035.6.3	Uncharacterised protein C1orf198	Q9H425	rs637914	18	32456099	T	A	0.67	0.67	trans	-	-
C1orf226.7989.5.3	Uncharacterised protein C1orf226	A1L170	rs11498715	10	82850327	T	G	0.86	0.86	trans	-	-
C6.4127.75.1	Complement component C6	P13671	rs4237676	11	39567316	C	T	0.42	0.44	trans	-	-
CA9.3798.71.1	Carbonic anhydrase 9	Q16790	rs1234192	2	229244950	C	T	0.53	0.54	trans	-	-
CACNA2D3.8885.6.3	Voltage-dependent calcium channel subunit alpha-2/delta-3	Q8IZS8	rs12866815	13	99760684	T	C	0.72	0.73	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
CACYBP.12432.23.3	Calcyclin-binding protein	Q9HB71	rs12626681	21	46814513	G	T	0.72	0.72	trans	-	-
CALB1.9918.23.3	Calbindin	P05937	rs1409177	13	94997242	A	C	0.96	0.96	trans	-	-
CALN1.10933.107.3	Calcium-binding protein 8	Q9BXU9	rs9327896	5	103486781	T	A	0.43	0.43	trans	-	-
CALR.5264.65.3	Calreticulin	P27797	rs117799470	13	54818489	C	T	0.97	0.97	trans	-	-
CALR.5264.65.3	Calreticulin	P27797	rs957327	15	33546275	A	G	0.47	0.47	trans	-	-
CAPN2.14684.17.3	Calpain-2 catalytic subunit	P17655	rs34685614	4	56150622	T	C	0.95	0.95	trans	-	-
CARHSP1.12808.103.3	Calcium-regulated heat stable protein 1	Q9Y2V2	rs11640887	16	86051882	C	T	0.71	0.70	trans	-	-
CARS.14098.28.3	Cysteine--tRNA ligase, cytoplasmic	P49589	rs553750511	22	48015355	T	C	0.97	0.97	trans	-	-
CARS.14098.28.3	Cysteine--tRNA ligase, cytoplasmic	P49589	rs4561701	2	68879540	A	G	0.73	0.73	trans	-	-
CASP10.5340.24.3	Caspase-10	Q92851	rs11835364	12	126284074	C	T	0.94	0.95	trans	-	-
CASS4.12855.16.3	Cas scaffolding protein family member 4	Q9NQ75	rs12694096	2	208834227	T	C	0.10	0.09	trans	-	-
CAT.3488.64.2	Catalase	P04040	rs62453446	7	39299749	T	A	0.12	0.11	trans	-	-
CBL.12016.60.3	E3 ubiquitin-protein ligase CBL	P22681	rs11088999	21	46811332	C	A	0.71	0.72	trans	-	-
CBLN1.9313.27.3	Cerebellin-1	P23435	rs17157815	7	29400704	T	C	0.96	0.96	trans	-	-
CBS.10086.39.3	Cystathionine beta-synthase	P35520	rs7312302	12	78013663	A	T	0.96	0.95	trans	-	-
CCBL2.12682.5.3	Kynurenine--oxoglutarate transaminase 3	Q6YP21	rs11195705	10	113506305	C	G	0.97	0.97	trans	-	-
CCL1.13687.5.3	C-C motif chemokine 1	P22362	rs113241089	12	132038318	G	C	0.97	0.96	trans	no	-
CCL1.13687.5.3	C-C motif chemokine 1	P22362	rs12424372	12	53037772	T	C	0.80	0.79	trans	no	-
CCL19.4922.13.1	C-C motif chemokine 19	Q99731	rs13188960	5	35853319	G	T	0.72	0.72	trans	-	-
CCL21.2516.57.3	C-C motif chemokine 21	O00585	rs73085438	20	13874204	C	T	0.91	0.92	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
CD177.13116.25.3	CD177 antigen	Q8N6Q3	rs17421832	1	66585962	C	T	0.87	0.87	trans	-	-
CD226.5721.1.3	CD226 antigen	Q15762	rs404449	7	4926950	T	C	0.61	0.63	trans	-	no
CD300A.5630.48.3	CMRF35-like molecule 8	Q9UGN4	rs3872610	3	112740145	C	T	0.18	0.19	trans	-	-
CD300E.10798.4.3	CMRF35-like molecule 2	Q496F6	rs6729654	2	51534914	A	T	0.28	0.28	trans	-	-
CD3G.7924.7.3	T-cell surface glycoprotein CD3 gamma chain	P09693	rs9571482	13	66383725	A	C	0.62	0.62	trans	-	-
CD4.3143.3.1	T-cell surface glycoprotein CD4	P01730	rs4811121	20	49714264	G	A	0.57	0.57	trans	-	-
CD44.9283.8.3	CD44 antigen	P16070	rs12901497	15	36623211	A	C	0.72	0.72	trans	-	-
CD5.14065.11.3	T-cell surface glycoprotein CD5	P06127	rs34169171	20	30194853	C	T	0.96	0.97	trans	-	-
CD5.14065.11.3	T-cell surface glycoprotein CD5	P06127	rs10469849	2	100318617	T	C	0.93	0.94	trans	-	-
CD58.10938.13.3	Lymphocyte function-associated antigen 3	P19256	rs8096899	18	58094046	G	T	0.95	0.95	trans	-	-
CD74.6974.6.3	HLA class II histocompatibility antigen gamma chain	P04233	rs4785468	16	50935200	T	C	0.46	0.45	trans	-	no
CD96.9735.44.3	T-cell surface protein tactile	P40200	rs893874	8	2010263	G	C	0.87	0.87	trans	-	-
CD97.2822.56.2	CD97 antigen	P48960	rs7953479	12	25518438	G	C	0.50	0.52	trans	-	-
CDH2.3797.1.1	Cadherin-2	P19022	rs1943603	18	37763127	T	A	0.83	0.83	trans	-	-
CDH2.3797.1.1	Cadherin-2	P19022	rs28421485	1	242145369	G	A	0.65	0.65	trans	-	-
CDKN1B.3719.2.2	Cyclin-dependent kinase inhibitor 1B	P46527	rs4819094	21	46812191	T	G	0.73	0.73	trans	-	-
CDON.4541.49.2	Cell adhesion molecule-related/down-regulated by oncogenes	Q4KMG0	rs140593096	18	48644235	C	T	0.97	0.97	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
CEACAM21.7204.1.3	Carcinoembryonic antigen-related cell adhesion molecule 21	Q3KPI0	rs78074431	11	115614360	G	A	0.95	0.94	trans	-	-
CELA2A.7140.1.3	Chymotrypsin-like elastase family member 2A	P08217	rs74366845	3	102161731	C	A	0.77	0.78	trans	-	-
CELA2A.7140.1.3	Chymotrypsin-like elastase family member 2A	P08217	rs9831862	3	146514079	C	T	0.58	0.57	trans	-	-
CFAP45.6553.68.3	Cilia- and flagella-associated protein 45	Q9UL16	rs10984054	9	121263967	T	C	0.91	0.91	trans	-	-
CFHR5.3666.17.4	Complement factor H-related protein 5	Q9BXR6	rs2474033	14	92392579	C	A	0.05	0.04	trans	no	-
CGA.FSHB.3032.11.2	Follicle stimulating hormone	P01215,P01225	rs6790495	3	40366770	C	T	0.67	0.70	trans	-	-
CGA.LHB.2953.31.2	Luteinizing hormone	P01215,P01229	rs78537284	19	49515363	G	A	0.92	0.92	cis	-	-
CGA.LHB.2953.31.2	Luteinizing hormone	P01215,P01229	rs139643250	19	49517146	C	T	0.92	0.93	cis	-	-
CGA.LHB.2953.31.2	Luteinizing hormone	P01215,P01229	rs12530875	7	106619472	G	A	0.93	0.93	trans	-	-
CGA.TSHB.3521.16.2	Thyroid Stimulating Hormone	P01215,P01222	rs112721893	16	531068	C	T	0.52	0.51	trans	-	-
CHCHD10.11270.17.3	Coiled-coil-helix-coiled-coil-helix domain-containing protein 10, mitochondrial	Q8WYQ3	rs146728922	10	69657144	T	A	0.96	0.96	trans	-	-
CHGB.8235.48.3	Secretogranin-1	P05060	rs10788525	10	88436490	C	T	0.54	0.53	trans	-	-
CHKB.13117.232.3	Choline/ethanolamine kinase	Q9Y259	rs6725784	2	39539427	G	A	0.97	0.96	trans	-	-
CHST15.14097.86.3	Carbohydrate sulfotransferase 15	Q7LFX5	rs11252810	10	4928428	C	T	0.92	0.93	trans	-	no
CHST3.7189.55.3	Carbohydrate sulfotransferase 3	Q7LGC8	rs58583902	9	110878023	T	A	0.96	0.96	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
CHST9.11646.4.3	Carbohydrate sulfotransferase 9	Q7L1S5	rs1434887	9	26199895	T	C	0.89	0.91	trans	-	-
CLEC2A.10953.14.3	C-type lectin domain family 2 member A	Q6UVW9	rs6876150	5	1028876	A	G	0.49	0.50	trans	-	-
CLEC7A.3603.60.5	C-type lectin domain family 7 member A	Q9BXN2	rs116862758	14	62005414	A	G	0.97	0.97	trans	-	-
CLEC7A.3603.60.5	C-type lectin domain family 7 member A	Q9BXN2	rs11716867	3	54630705	C	T	0.71	0.68	trans	-	-
CLIC5.12475.48.3	Chloride intracellular channel protein 5	Q9NZA1	rs11981812	7	154137194	G	A	0.77	0.77	trans	-	-
CLINT1.11659.31.3	Clathrin interactor 1	Q14677	rs113158475	11	69580923	C	T	0.92	0.93	trans	-	-
CLMP.10440.26.3	CXADR-like membrane protein	Q9H6B4	rs911148	20	58214282	C	A	0.31	0.32	trans	no	-
CLN5.9795.9.3	Ceroid-lipofuscinosis neuronal protein 5	O75503	rs566185205	12	61030703	A	C	0.89	0.88	trans	no	-
CLPS.5749.53.3	Colipase	P04118	rs242057	1	4773852	G	A	0.26	0.26	trans	-	-
CMA1.3423.59.2	Chymase	P23946	rs11580550	1	106031366	A	G	0.80	0.80	trans	-	-
CMPK1.8312.139.3	UMP-CMP kinase	P30085	rs2248409	13	80320345	A	G	0.43	0.45	trans	no	-
COA3.7888.58.3	Cytochrome c oxidase assembly factor 3 homolog, mitochondrial	Q9Y2R0	rs9974464	21	40500901	C	T	0.16	0.17	trans	-	-
COG8.9543.131.3	Conserved oligomeric Golgi complex subunit 8	Q96MW5	rs11216799	11	118046570	G	A	0.91	0.90	trans	-	-
COL11A2.11278.4.3	Collagen alpha-2(XI) chain	P13942	rs148930599	5	150862554	C	T	0.89	0.89	trans	-	-
COL1A1.11140.56.3	Collagen alpha-1(I) chain	P02452	rs3123687	10	31529261	C	G	0.06	0.06	trans	-	-
COL1A1.11140.56.3	Collagen alpha-1(I) chain	P02452	rs114160127	1	231219360	G	A	0.97	0.97	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
COL20A1.8804.39.3	Collagen alpha-1(XX) chain	Q9P218	rs12442553	15	98907097	C	T	0.97	0.97	trans	-	-
COL6A1.11150.3.3	Collagen alpha-1(VI) chain	P12109	rs150370694	10	27258842	C	T	0.96	0.96	trans	-	-
COL8A1.4807.13.3	Collagen alpha-1(VIII) chain	P27658	rs113936403	5	178586315	G	A	0.97	0.97	trans	-	-
COMMD7.2823.7.1	COMM domain-containing protein 7	Q86VX2	rs77341824	18	6474441	T	C	0.89	0.87	trans	-	-
COTL1.4905.63.1	Coactosin-like protein	Q14019	rs34592426	6	31239090	G	C	0.82	0.81	trans	-	-
COX5B.7887.57.3	Cytochrome c oxidase subunit 5B, mitochondrial	P10606	rs2884815	12	62366581	T	C	0.60	0.59	trans	-	-
CPB1.6356.3.3	Carboxypeptidase B	P15086	rs78040619	2	154823175	C	A	0.96	0.96	trans	-	-
CPVL.8891.7.3	Probable serine carboxypeptidase CPVL	Q9H3G5	rs2686200	8	11664540	A	T	0.63	0.62	trans	-	-
CPXM1.6255.74.3	Probable carboxypeptidase X1	Q96SM3	rs116928246	12	125025742	G	A	0.54	0.55	trans	-	-
CPZ.6493.9.3	Carboxypeptidase Z	Q66K79	rs11099939	4	155078353	G	A	0.03	0.03	trans	-	-
CROT.13929.27.3	Peroxisomal carnitine O-octanoyltransferase	Q9UKG9	rs36113849	18	22801268	T	C	0.93	0.94	trans	-	-
CRYBB2.10000.28.3	Beta-crystallin B2	P43320	rs2282160	9	6534845	T	C	0.56	0.54	trans	-	-
CSPG4.8951.162.3	Chondroitin sulfate proteoglycan 4	Q6UVK1	rs7929804	11	2640824	G	A	0.46	0.46	trans	-	-
CTNNA2.9872.23.3	Catenin alpha-2	P26232	rs10738148	9	1018245	C	T	0.37	0.35	trans	-	-
CTSF.9212.22.3	Cathepsin F	Q9UBX1	rs36012476	19	41206060	G	C	0.95	0.96	trans	no	-
CTSV.3364.76.2	Cathepsin L2	O60911	rs62512308	8	59193808	G	T	0.96	0.96	trans	-	-
CTSV.3364.76.2	Cathepsin L2	O60911	rs77290088	9	101122455	G	T	0.97	0.97	trans	-	-
CUL3.10045.47.3	Cullin-3	Q13618	rs215372	12	47745802	A	G	0.68	0.67	trans	-	-
CUL4B.13743.56.3	Cullin-4B	Q13620	rs10993060	9	96887225	G	A	0.84	0.85	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
CXCL11.3038.9.2	C-X-C motif chemokine 11	O14625	rs35170645	4	76964956	A	G	0.49	0.48	cis	-	-
CXCL12.2330.2.1	Stromal cell-derived factor 1	P48061	rs28604315	14	106935695	A	C	0.68	0.69	trans	-	-
CXCL13.3487.32.2	C-X-C motif chemokine 13	O43927	rs117939745	10	61344734	C	A	0.96	0.96	trans	-	-
CXCL16.2436.49.4	C-X-C motif chemokine 16	Q9H2A7	rs71586562	5	104335989	C	T	0.71	0.69	trans	-	-
CXCL5.2979.8.2	C-X-C motif chemokine 5	P42830	rs9853250	3	59183655	C	G	0.45	0.46	trans	-	-
CXCL9.11593.21.3	C-X-C motif chemokine 9	Q07325	rs2470174	15	51670821	C	T	0.91	0.91	trans	-	no
CYGB.11546.7.3	Cytoglobin	Q8WWM9	rs75067265	12	112993771	C	G	0.96	0.96	trans	-	-
DCAF5.11283.13.3	DDB1- and CUL4-associated factor 5	Q96JK2	rs1428250	5	38216412	G	T	0.04	0.04	trans	-	-
DCN.2666.53.2	Decorin	P07585	rs2515031	8	119490850	G	C	0.60	0.60	trans	-	-
DCP1A.14008.22.3	mRNA-decapping enzyme 1A	Q9NPI6	rs72837699	17	7112250	C	G	0.92	0.91	trans	-	-
DCUN1D3.13553.4.3	DCN1-like protein 3	Q8IWE4	rs35269401	2	103536438	G	C	0.86	0.87	trans	-	-
DCUN1D3.13553.4.3	DCN1-like protein 3	Q8IWE4	rs4524627	6	123992030	A	G	0.88	0.88	trans	-	-
DDX39B.9742.59.3	Spliceosome RNA helicase DDX39B	Q13838	rs56292770	21	46810010	G	A	0.71	0.72	trans	-	-
DDX39B.9742.59.3	Spliceosome RNA helicase DDX39B	Q13838	rs2023222	6	32194083	T	G	0.63	0.63	trans	-	-
DGCR6.9444.70.3	Protein DGCR6	Q14129	rs7024075	9	11274869	C	A	0.72	0.71	trans	-	-
DHFR.9823.2.3	Dihydrofolate reductase	P00374	rs6910044	6	139852731	A	C	0.63	0.63	trans	-	-
DHFR.9823.2.3	Dihydrofolate reductase	P00374	rs55957950	8	16192954	A	G	0.71	0.72	trans	-	-
DIABLO.3122.6.2	Diablo homolog, mitochondrial	Q9NR28	rs10250463	7	127693537	G	A	0.86	0.87	trans	-	-
DLL1.5349.69.3	Delta-like protein 1	O00548	rs9295988	6	31373718	C	G	0.93	0.93	trans	-	no

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
DNAJB9.11214.40.3	DnaJ homolog subfamily B member 9	Q9UBS3	rs76855778	1	239908310	A	G	0.96	0.96	trans	-	-
DNAJC17.14655.1.3	DnaJ homolog subfamily C member 17	Q9NVM6	rs58914460	19	48350847	C	T	0.79	0.79	trans	-	-
DNAJC18.8033.1.3	DnaJ homolog subfamily C member 18	Q9H819	rs9841975	3	64446112	A	G	0.54	0.55	trans	-	-
DNAJC4.8016.19.3	DnaJ homolog subfamily C member 4	Q9NNZ3	rs703473	19	55458816	G	A	0.96	0.95	trans	-	-
DNAJC4.8016.19.3	DnaJ homolog subfamily C member 4	Q9NNZ3	rs4858378	3	22154087	A	G	0.70	0.70	trans	-	-
DNER.11442.1.3	Delta and Notch-like epidermal growth factor-related receptor	Q8NFT8	rs7249288	19	53601493	T	C	0.82	0.82	trans	-	no
DPEP1.8794.13.3	Dipeptidase 1	P16444	rs59423625	2	182493374	C	A	0.92	0.92	trans	-	-
DPYSL3.12707.26.3	Dihydropyrimidinase-related protein 3	Q14195	rs11646557	16	86042491	A	G	0.74	0.73	trans	-	-
DRGX.8034.6.3	Dorsal root ganglia homeobox protein	A6NNA5	rs75678034	10	79214065	C	T	0.92	0.92	trans	-	-
DSC2.13126.52.3	Desmocollin-2	Q02487	rs533379	13	51172579	G	C	0.34	0.33	trans	-	-
DSC3.4981.6.1	Desmocollin-3	Q14574	rs2428085	9	136871504	C	T	0.08	0.08	trans	-	-
DSG1.2976.58.2	Desmoglein-1	Q02413	rs113823715	1	219863519	T	C	0.92	0.91	trans	-	-
DUSP4.10035.6.3	Dual specificity protein phosphatase 4	Q13115	rs2515797	11	117818680	G	A	0.41	0.40	trans	-	-
ECE1.3611.70.4	Endothelin-converting enzyme 1	P42892	rs116386809	1	208484335	G	A	0.97	0.98	trans	-	no
ECE1.8767.44.3	Endothelin-converting enzyme 1	P42892	rs11047985	12	25529645	C	T	0.50	0.48	trans	-	no
EFNA3.5759.10.3	Ephrin-A3	P52797	rs58769948	15	98629996	G	A	0.88	0.89	trans	-	-
EFNB3.2514.65.3	Ephrin-B3	Q15768	rs17065165	4	178678401	A	G	0.74	0.73	trans	-	no

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
EGFR.5328.33.37	Epidermal growth factor receptor variant III	P00533	rs10468843	18	33995541	G	C	0.78	0.79	trans	-	-
EIF3G.11454.87.3	Eukaryotic translation initiation factor 3 subunit G	O75821	rs6565994	18	75241668	G	C	0.28	0.27	trans	-	-
EIF4E2.9745.20.3	Eukaryotic translation initiation factor 4E type 2	O60573	rs6576171	14	106923019	A	G	0.75	0.75	trans	-	-
ELK1.10006.25.3	ETS domain-containing protein Elk-1	P19419	rs142806824	16	79276552	A	G	0.44	0.43	trans	-	-
ELMO1.12764.3.3	Engulfment and cell motility protein 1	Q92556	rs828842	2	74355244	G	A	0.56	0.58	trans	-	-
ENO2.10339.48.3	Gamma-enolase	P09104	rs34428701	10	20634377	T	C	0.89	0.88	trans	no	-
EPHA1.3431.54.2	Ephrin type-A receptor 1	P21709	rs78237190	1	12831095	G	A	0.96	0.96	trans	-	-
EPHA2.4834.61.2	Ephrin type-A receptor 2	P29317	rs911148	20	58214282	C	A	0.31	0.32	trans	-	-
ERAP1.4964.67.1	Endoplasmic reticulum aminopeptidase 1	Q9NZ08	rs6859054	5	158903001	T	G	0.81	0.82	trans	-	-
ERCC4.9895.77.3	DNA repair endonuclease XPF	Q92889	rs113627905	3	195226171	C	T	0.56	0.55	trans	-	-
ERP29.13728.19.3	Endoplasmic reticulum resident protein 29	P30040	rs10950846	7	21401732	A	G	0.92	0.93	trans	-	-
ERP44.6064.4.3	Endoplasmic reticulum resident protein 44	Q9BS26	rs6074977	20	15931275	C	A	0.03	0.03	trans	-	-
ERP44.6064.4.3	Endoplasmic reticulum resident protein 44	Q9BS26	rs1012334	21	34713751	A	T	0.53	0.54	trans	-	-
ESAM.2981.9.3	Endothelial cell-selective adhesion molecule	Q96AP7	rs117400722	6	102449616	G	T	0.96	0.96	trans	-	no
ESD.4984.83.1	S-formylglutathione hydrolase	P10768	rs4646391	17	17441369	C	G	0.46	0.47	trans	-	-
ETS2.12350.86.3	Protein C-ets-2	P15036	rs57192414	4	162200517	T	C	0.72	0.73	trans	-	-
EVA1C.8877.22.3	Protein eva-1 homolog C	P58658	rs1983765	9	3151278	T	C	0.63	0.62	trans	no	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
FAIM3.6574.11.3	Fas apoptotic inhibitory molecule 3	O60667	rs61940590	12	129148029	G	A	0.47	0.48	trans	-	-
FAM134B.8556.5.3	Protein FAM134B	Q9H6L5	rs2983713	1	45168726	T	A	0.83	0.84	trans	-	-
FAM150B.6284.7.3	Protein FAM150B	Q6UX46	rs17289966	5	58216642	T	C	0.92	0.92	trans	-	-
FAM151A.7856.51.3	Protein FAM151A	Q8WW52	rs582835	6	94728619	T	C	0.33	0.34	trans	-	-
FAM163B.10880.38.3	Protein FAM163B	P0C2L3	rs10942276	5	71640481	G	A	0.95	0.95	trans	-	-
FAM171B.8061.102.3	Protein FAM171B	Q6P995	rs9472777	6	12872092	G	A	0.74	0.74	trans	-	no
FAM171B.8786.6.3	Protein FAM171B	Q6P995	rs114439777	5	147245561	C	T	0.79	0.79	trans	-	no
FAM171B.8851.42.3	Protein FAM171B	Q6P995	rs79585054	1	1414689	C	T	0.94	0.94	trans	no	-
FAM171B.8851.42.3	Protein FAM171B	Q6P995	rs117153691	9	20523966	C	G	0.95	0.95	trans	-	no
FAM174A.6597.24.3	Membrane protein FAM174A	Q8TBP5	rs4485089	11	12431361	A	G	0.74	0.73	trans	-	-
FAM174A.6597.24.3	Membrane protein FAM174A	Q8TBP5	rs2723264	12	40652520	T	C	0.19	0.19	trans	-	-
FAM19A4.6511.17.3	Protein FAM19A4	Q96LR4	rs12500806	4	69475763	C	T	0.37	0.35	trans	-	-
FAM213A.13423.94.3	Redox-regulatory protein FAM213A	Q9BRX8	rs7862702	9	8748282	C	T	0.31	0.30	trans	-	-
FAM24B.8775.61.3	Protein FAM24B	Q8N5W8	rs4938846	11	56664501	C	T	0.89	0.89	trans	-	-
FAS.9459.7.3	Tumor necrosis factor receptor superfamily member 6	P25445	rs263332	8	5576744	T	C	0.76	0.77	trans	-	no
FBLN1.6470.19.3	Fibulin-1	P23142	rs6552526	4	182751663	C	T	0.08	0.08	trans	-	-
FBXO3.14628.72.3	F-box only protein 3	Q9UK99	rs1394564	16	10677181	A	G	0.35	0.36	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
FCGR2A.3309.2.2	Low affinity immunoglobulin gamma Fc region receptor II-a	P12318	rs73193587	3	192115823	T	A	0.66	0.66	trans	-	-
FCGR2A.3309.2.2	Low affinity immunoglobulin gamma Fc region receptor II-a	P12318	rs6900370	6	112663749	A	G	0.89	0.89	trans	-	-
FCRL5.6103.70.3	Fc receptor-like protein 5	Q96RD9	rs115230181	1	31423361	A	G	0.98	0.98	trans	-	-
FCRL6.6617.12.3	Fc receptor-like protein 6	Q6DN72	rs7913390	10	24906143	C	T	0.69	0.69	trans	-	-
FGF4.4123.60.2	Fibroblast growth factor 4	P08620	rs1031267	2	6428941	T	C	0.10	0.10	trans	-	-
FGF7.4487.1.1	Fibroblast growth factor 7	P21781	rs10146530	14	89461522	C	T	0.74	0.71	trans	-	no
FGF7.4487.1.1	Fibroblast growth factor 7	P21781	rs150021587	1	242206422	G	A	0.97	0.96	trans	no	-
FGFR4.4988.49.2	Fibroblast growth factor receptor 4	P22455	rs59231956	4	4872748	C	A	0.86	0.85	trans	-	-
FKBP7.9288.7.3	Peptidyl-prolyl cis-trans isomerase FKBP7	Q9Y680	rs3092288	20	43507848	G	A	0.73	0.72	trans	-	-
FKBP7.9288.7.3	Peptidyl-prolyl cis-trans isomerase FKBP7	Q9Y680	rs28670583	9	17937867	A	G	0.62	0.62	trans	-	-
FLRT3.9128.34.3	Leucine-rich repeat transmembrane protein FLRT3	Q9NZU0	rs34145303	18	72420305	A	G	0.94	0.94	trans	-	no
FLRT3.9128.34.3	Leucine-rich repeat transmembrane protein FLRT3	Q9NZU0	rs882275	2	129125972	C	A	0.48	0.47	trans	-	no
FOXJ2.14204.55.3	Forkhead box protein J2	Q9P0K8	rs138184725	8	5955964	G	A	0.97	0.97	trans	-	-
FOXL2.11375.49.3	Forkhead box protein L2	P58012	rs1411381	9	27707461	G	T	0.72	0.71	trans	-	-
FST.4132.27.2	Follistatin	P19883	rs56359796	7	149697994	G	A	0.39	0.39	trans	-	-
FTH1.FTL.3031.66.1	Ferritin	P02792,P02794	rs75632028	16	8389374	A	C	0.89	0.89	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
FUT3.4548.4.2	Galactoside 3(4)-L-fucosyltransferase	P21217	rs12456689	18	56859179	T	A	0.80	0.82	trans	-	-
FUT9.6991.24.3	Alpha-(1,3)-fucosyltransferase 9	Q9Y231	rs336962	5	82971347	A	T	0.42	0.42	trans	-	-
G6B.8659.68.3	Protein G6b	O95866	rs903836	11	10852822	T	C	0.03	0.04	trans	-	no
GBP6.7818.101.3	Guanylate-binding protein 6	Q6ZN66	rs10077981	5	37901492	A	G	0.51	0.50	trans	-	-
GCK.12960.9.3	Glucokinase	P35557	rs12708701	16	10584311	T	G	0.87	0.89	trans	-	-
GCK.12960.9.3	Glucokinase	P35557	rs573520313	1	12887220	T	G	0.95	0.95	trans	-	-
GFRA2.2515.14.3	GDNF family receptor alpha-2	O00451	rs116025596	3	117210567	C	T	0.97	0.97	trans	-	-
GFRA2.2515.14.3	GDNF family receptor alpha-2	O00451	rs112216801	6	32601838	C	A	0.76	0.77	trans	-	-
GLIPR1.9265.10.3	Glioma pathogenesis-related protein 1	P48060	rs7806154	7	8256455	T	C	0.86	0.84	trans	-	-
GLTP.12513.8.3	Glycolipid transfer protein	Q9NZD2	rs2718778	3	180944824	G	A	0.14	0.13	trans	-	-
GNPNAT1.13954.9.3	Glucosamine 6-phosphate N-acetyltransferase	Q96EK6	rs80334360	7	25675528	A	G	0.88	0.89	trans	-	-
GNPTG.10666.7.3	N-acetylglucosamine-1-phosphotransferase subunit gamma	Q9UJJ9	rs10066049	5	88434324	T	A	0.81	0.80	trans	-	-
GNS.8314.71.3	N-acetylglucosamine-6-sulfatase	P15586	rs7717453	5	179617421	C	T	0.60	0.58	trans	-	no
GPD1L.12420.10.3	Glycerol-3-phosphate dehydrogenase 1-like protein	Q8N335	rs10077981	5	37901492	A	G	0.51	0.50	trans	-	-
GPI.4272.46.2	Glucose-6-phosphate isomerase	P06744	rs11088998	21	46810776	G	A	0.76	0.77	trans	-	-
GPI.4272.46.2	Glucose-6-phosphate isomerase	P06744	rs77546608	6	18970054	A	T	0.97	0.96	trans	-	-
GPNMB.5080.131.3	Transmembrane glycoprotein NMB	Q14956	rs72702815	9	10555820	T	C	0.96	0.95	trans	no	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
GRAMD1C.8842.16.3	GRAM domain-containing protein 1C	Q8IYS0	rs61077924	3	113625933	C	G	0.68	0.68	cis	yes	yes
GRAMD1C.8842.16.3	GRAM domain-containing protein 1C	Q8IYS0	rs4682146	3	113626645	C	G	0.67	0.67	cis	yes	yes
GRP.5897.58.3	Gastrin-releasing peptide	P07492	rs4817013	21	26608365	T	C	0.58	0.57	trans	no	-
GRP.5897.58.3	Gastrin-releasing peptide	P07492	rs4872232	8	24279067	C	A	0.43	0.42	trans	no	-
GRP.8400.74.3	Gastrin-releasing peptide	P07492	rs72892585	18	25934326	C	A	0.97	0.97	trans	-	no
GSTA4.14645.253.3	Glutathione S-transferase A4	O15217	rs33999362	3	53571914	A	G	0.58	0.57	trans	-	-
GSTK1.13474.40.3	Glutathione S-transferase kappa 1	Q9Y2Q3	rs62176181	2	166052014	G	T	0.90	0.90	trans	-	-
GSTK1.13474.40.3	Glutathione S-transferase kappa 1	Q9Y2Q3	rs112603185	4	169466549	A	G	0.96	0.96	trans	-	-
GUCA2B.6223.5.3	Guanylate cyclase activator 2B	Q16661	rs10887641	10	88431738	A	G	0.50	0.49	trans	-	-
GZMH.3373.5.2	Granzyme H	P20718	rs730943	20	46795863	A	T	0.68	0.67	trans	-	-
H2AFZ.4163.5.2	Histone H2A.z	P0C0S5	rs7448318	5	63252922	C	T	0.31	0.32	trans	-	-
HAVCR1.9021.1.3	Hepatitis A virus cellular receptor 1	Q96D42	rs12490183	3	153274971	C	T	0.94	0.94	trans	-	-
HBEGF.14094.29.3	Heparin-binding EGF-like growth factor	Q99075	rs12490098	3	45232261	A	T	0.95	0.96	trans	-	no
HDGFRP2.4553.65.3	Hepatoma-derived growth factor-related protein 2	Q7Z4V5	rs56175853	6	96149396	A	G	0.85	0.85	trans	-	-
HECW1.12669.30.3	E3 ubiquitin-protein ligase HECW1	Q76N89	rs17636165	5	3540085	G	A	0.82	0.80	trans	-	-
HEPACAM2.6568.18.3	HEPACAM family member 2	A8MVW5	rs73133989	3	98410273	C	A	0.95	0.96	trans	-	no
HEPACAM2.9116.28.3	HEPACAM family member 2	A8MVW5	rs160283	5	82859485	G	A	0.75	0.76	trans	-	no

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
HEPACAM2.9116.28.3	HEPACAM family member 2	A8MVW5	rs13156973	5	99943991	A	G	0.57	0.59	trans	no	-
HEPACAM2.9116.28.3	HEPACAM family member 2	A8MVW5	rs75311297	8	11736582	G	A	0.92	0.93	trans	no	-
HEPACAM2.9116.28.3	HEPACAM family member 2	A8MVW5	rs77451450	9	18432802	T	C	0.88	0.88	trans	-	no
HMGB3.12775.6.3	High mobility group protein B3	O15347	rs344956	3	187895427	C	A	0.47	0.48	trans	-	-
HNRNPDL.10852.114.3	Heterogeneous nuclear ribonucleoprotein D-like	O14979	rs11088999	21	46811332	C	A	0.71	0.72	trans	-	-
HNRNPM.12783.29.3	Heterogeneous nuclear ribonucleoprotein M	P52272	rs56292770	21	46810010	G	A	0.71	0.72	trans	-	-
HNRNPM.12783.29.3	Heterogeneous nuclear ribonucleoprotein M	P52272	rs3132950	6	32194740	C	T	0.63	0.63	trans	-	-
HOGA1.10024.44.3	4-hydroxy-2-oxoglutarate aldolase, mitochondrial	Q86XE5	rs29579	16	8383753	C	G	0.80	0.79	trans	-	-
HRK.8008.28.3	Activator of apoptosis harakiri	O00198	rs7277498	21	21030126	A	G	0.72	0.71	trans	-	-
HS3ST3B1.6986.17.3	Heparan sulfate glucosamine 3-O-sulfotransferase 3B1	Q9Y662	rs56311947	1	108360297	G	A	0.62	0.62	trans	-	-
HTRA2.3317.33.1	Serine protease HTRA2, mitochondrial	O43464	rs4306167	1	64637880	G	T	0.69	0.69	trans	-	-
ICAM5.8245.27.3	Intercellular adhesion molecule 5	Q9UMF0	rs78706926	14	105143191	T	C	0.89	0.90	trans	no	-
ICOSLG.9303.9.3	ICOS ligand	O75144	rs115053672	1	69375691	C	T	0.97	0.96	trans	-	no
ID1.9436.2.3	DNA-binding protein inhibitor ID-1	P41134	rs680119	11	85716032	T	C	0.39	0.40	trans	-	-
IDE.3197.70.2	Insulin-degrading enzyme	P14735	rs13211315	6	143255891	G	C	0.64	0.64	trans	-	-
IDO1.9759.13.3	Indoleamine 2,3-dioxygenase 1	P14902	rs12310980	12	131220472	G	A	0.66	0.66	trans	-	-
IFNA14.7180.114.3	Interferon alpha-14	P01570	rs17006826	4	84370383	T	C	0.63	0.63	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
IFNG.14147.50.3	Interferon gamma	P01579	rs9295360	6	167089915	T	C	0.60	0.61	trans	-	no
IFNL1.13734.22.3	Interferon lambda-1	Q8IU54	rs10067159	5	68250691	C	T	0.47	0.47	trans	-	-
IGDCC3.11952.1.3	Immunoglobulin superfamily DCC subclass member 3	Q8IVU1	rs9267845	6	32193698	A	T	0.63	0.63	trans	-	no
IGFBP1.13741.36.3	Insulin-like growth factor-binding protein 1	P08833	rs2050087	20	40572429	C	T	0.78	0.80	trans	-	-
IGFBPL1.7815.49.3	Insulin-like growth factor-binding protein-like 1	Q8WX77	rs2662723	4	121236664	G	C	0.44	0.44	trans	-	-
IGHE.IGK.IGL.4135.84.2	Immunoglobulin E	P01854	rs752626	1	15602330	C	T	0.75	0.75	trans	-	-
IGJ.8362.102.3	Immunoglobulin J chain	P01591	rs72769720	9	125110968	C	T	0.94	0.94	trans	-	-
IL12A.IL12B.10367.26.3	Interleukin-12	P29459,P29460	rs3851630	12	104135239	T	C	0.18	0.18	trans	-	no
IL12A.IL12B.10367.26.3	Interleukin-12	P29459,P29460	rs79791086	14	87607863	T	C	0.97	0.97	trans	no	-
IL15RA.3445.53.2	Interleukin-15 receptor subunit alpha	Q13261	rs9599203	13	67811553	A	C	0.90	0.90	trans	no	-
IL17C.9255.5.3	Interleukin-17C	Q9P0M4	rs144631133	3	154338234	A	G	0.96	0.96	trans	-	-
IL17RB.5084.154.3	Interleukin-17 receptor B	Q9NRM6	rs77301136	11	132764647	C	G	0.96	0.96	trans	-	-
IL17RB.5084.154.3	Interleukin-17 receptor B	Q9NRM6	rs75029903	5	65764134	C	T	0.93	0.94	trans	-	-
IL17RC.5468.67.3	Interleukin-17 receptor C	Q8NAC3	rs2267266	22	34105634	C	T	0.53	0.53	trans	-	-
IL17RD.3376.49.2	Interleukin-17 receptor D	Q8NFM7	rs35706962	11	106402698	C	T	0.92	0.93	trans	-	-
IL18R1.14079.14.3	Interleukin-18 receptor 1	Q13478	rs892350	8	130568904	T	C	0.76	0.75	trans	-	no
IL1R2.5666.64.3	Interleukin-1 receptor type 2	P27930	rs12676148	8	16734080	C	T	0.91	0.91	trans	-	no
IL22.2778.10.2	Interleukin-22	Q9GZX6	rs1575962	9	22282452	A	C	0.21	0.20	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
IL22RA1.9603.9.3	Interleukin-22 receptor subunit alpha-1	Q8N6P7	rs3128390	7	112085342	C	T	0.58	0.56	trans	-	no
IL22RA2.5087.5.3	Interleukin-22 receptor subunit alpha-2	Q969J5	rs78682114	4	89299518	C	T	0.93	0.94	trans	-	no
IL23R.5088.175.3	Interleukin-23 receptor	Q5VWK5	rs7449190	5	1297918	T	C	0.15	0.15	trans	-	-
IL25.4137.57.2	Interleukin-25	Q9H293	rs11103845	9	138252603	G	A	0.94	0.94	trans	-	-
IL34.4556.10.2	Interleukin-34	Q6ZMJ4	rs17418632	15	48106921	A	G	0.85	0.85	trans	-	-
IL5.3741.4.3	Interleukin-5	P05113	rs10981641	9	115900126	C	T	0.94	0.94	trans	-	no
ILF3.12759.47.3	Interleukin enhancer-binding factor 3	Q12906	rs7828742	8	116960729	A	G	0.40	0.38	trans	-	-
ILF3.12759.47.3	Interleukin enhancer-binding factor 3	Q12906	rs1898489	8	52206528	T	C	0.08	0.08	trans	-	-
IMPAD1.9231.23.3	Inositol monophosphatase 3	Q9NX62	rs11032153	11	33153026	G	A	0.95	0.95	trans	-	-
INSL3.5723.4.3	Insulin-like 3	P51460	rs59037764	11	131269982	G	T	0.72	0.72	trans	-	-
IRF1.10351.51.3	Interferon regulatory factor 1	P10914	rs73173432	22	48993735	G	A	0.89	0.88	trans	no	-
ISG15.14151.4.3	Ubiquitin-like protein ISG15	P05161	rs80038180	15	88984007	C	T	0.97	0.96	trans	no	-
ITGAL.11617.1.3	Integrin alpha-L	P20701	rs3928116	9	130766948	T	G	0.75	0.75	trans	-	-
ITIH1.7955.195.3	Inter-alpha-trypsin inhibitor heavy chain H1	P19827	rs9266806	6	31378022	T	C	0.79	0.80	trans	-	-
ITM2A.7765.15.3	Integral membrane protein 2A	O43736	rs543076146	5	95263384	C	T	0.81	0.81	trans	-	-
ITPRIPL1.9221.6.3	Inositol 1,4,5-trisphosphate receptor-interacting protein-like 1	Q6GPH6	rs56292770	21	46810010	G	A	0.71	0.72	trans	-	-
ITSN1.14070.56.3	Intersectin-1	Q15811	rs62580764	9	113027797	G	C	0.91	0.89	trans	-	-
JAK2.4998.50.2	Tyrosine-protein kinase JAK2	O60674	rs113065532	14	22015047	C	T	0.94	0.93	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
JAM2.2997.8.1	Junctional adhesion molecule B	P57087	rs10182134	2	142191945	T	C	0.51	0.52	trans	-	-
KAL1.6603.18.3	Anosmin-1	P23352	rs28733769	7	50419792	G	T	0.85	0.85	trans	-	-
KCNE3.8067.21.3	Potassium voltage-gated channel subfamily E member 3	Q9Y6H6	rs35652435	1	231161088	G	A	0.73	0.70	trans	-	-
KCNIP1.13650.11.3	Kv channel-interacting protein 1	Q9NZI2	rs2337632	12	123736345	A	G	0.36	0.35	trans	-	-
KIAA1161.8093.13.3	Uncharacterised family 31 glucosidase KIAA1161	Q6NSJ0	rs16823123	3	153466396	A	G	0.96	0.96	trans	-	no
KIAA1324.10637.50.3	UPF0577 protein KIAA1324	Q6UXG2	rs4424962	18	13575712	C	T	0.80	0.79	trans	no	-
KIAA2013.6538.90.3	Uncharacterised protein KIAA2013	Q8IYS2	rs79181095	1	161176519	G	A	0.91	0.90	trans	-	-
KIF16B.11672.17.3	Kinesin-like protein KIF16B	Q96L93	rs146761139	10	103805658	C	T	0.94	0.94	trans	-	-
KIR2DS2.10428.1.3	Killer cell immunoglobulin-like receptor 2DS2	P43631	rs28886328	4	138400096	T	C	0.93	0.92	trans	-	-
KIR3DS1.5097.14.3	Killer cell immunoglobulin-like receptor 3DS1	Q14943	rs1607567	5	22997098	G	T	0.46	0.45	trans	-	-
KIRREL3.4557.61.2	Kin of IRRE-like protein 3	Q8IZU9	rs77274954	11	122919030	A	C	0.95	0.96	trans	-	-
KITLG.9377.25.3	Kit ligand	P21583	rs2165346	16	60340509	T	C	0.80	0.80	trans	-	-
KLK14.3681.87.3	Kallikrein-14	Q9P0G3	rs11943860	4	37991472	G	T	0.72	0.71	trans	-	no
KLK14.8620.56.3	Kallikrein-14	Q9P0G3	rs11059313	12	128157614	T	C	0.49	0.51	trans	no	-
KLK3.8468.19.3	Prostate-specific antigen	P07288	rs10993994	10	51549496	T	C	0.40	0.40	trans	-	-
KLK3.8468.19.3	Prostate-specific antigen	P07288	rs151224625	16	69721169	C	G	0.97	0.97	trans	-	-
KLRC1.5629.58.3	NKG2-A/NKG2-B type II integral membrane protein	P26715	rs67373475	5	53151795	C	A	0.81	0.81	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
KLRC1.5629.58.3	NKG2-A/NKG2-B type II integral membrane protein	P26715	rs6947549	7	70996732	T	G	0.14	0.14	trans	-	-
KMT2C.11402.17.3	Histone-lysine N-methyltransferase 2C	Q8NEZ4	rs79892278	22	20219915	G	A	0.92	0.92	trans	-	-
KMT2C.11402.17.3	Histone-lysine N-methyltransferase 2C	Q8NEZ4	rs117992328	6	106364158	C	A	0.97	0.97	trans	-	-
KPNA2.2860.19.2	Importin subunit alpha-1	P52292	rs62149015	2	83427196	G	A	0.91	0.92	trans	-	-
KPNA4.12698.72.3	Importin subunit alpha-3	O00629	rs7777594	7	29637619	G	A	0.96	0.96	trans	-	-
KREMEN1.6512.68.3	Kremen protein 1	Q96MU8	rs35221840	17	71546115	G	A	0.95	0.95	trans	-	-
KRTAP2.4.14615.46.3	Keratin-associated protein 2-4	Q9BYR9	rs317198	11	89083844	G	A	0.91	0.92	trans	-	-
LAG3.5099.14.3	Lymphocyte activation gene 3 protein	P18627	rs2431501	5	176606052	T	A	0.36	0.35	trans	no	-
LAG3.9950.229.3	Lymphocyte activation gene 3 protein	P18627	rs28437271	3	66019291	C	T	0.96	0.96	trans	no	-
LAIR1.11284.24.3	Leukocyte-associated immunoglobulin-like receptor 1	Q6GTX8	rs793715	6	24445131	T	C	0.04	0.05	trans	-	-
LAMC2.9580.5.3	Laminin subunit gamma-2	Q13753	rs112783675	2	141412248	G	A	0.94	0.94	trans	-	-
LAT2.5613.75.3	Linker for activation of T-cells family member 2	Q9GZY6	rs2277329	12	53468419	T	G	0.78	0.76	trans	-	-
LCN2.2836.68.2	Neutrophil gelatinase-associated lipocalin	P80188	rs6871004	5	177978957	G	A	0.89	0.90	trans	-	-
LCT.9017.58.3	Lactase-phlorizin hydrolase	P09848	rs55638935	16	76632804	C	T	0.88	0.88	trans	-	-
LDLR.13129.40.3	Low-density lipoprotein receptor	P01130	rs58124292	16	81890435	G	A	0.80	0.80	trans	-	-
LDLR.13129.40.3	Low-density lipoprotein receptor	P01130	rs58096832	6	19153122	G	A	0.85	0.85	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
LECT1.9928.125.3	Leukocyte cell-derived chemotaxin 1	O75829	rs12620171	2	56480069	A	T	0.75	0.75	trans	-	-
LEP.8484.24.3	Leptin	P41159	rs17047302	2	56112675	G	A	0.90	0.91	trans	no	-
LGALS8.4909.68.1	Galectin-8	O00214	rs9587769	13	89191518	G	A	0.81	0.81	trans	-	-
LGALS8.4909.68.1	Galectin-8	O00214	rs6103273	20	42019919	G	A	0.38	0.37	trans	-	-
LGALS8.4909.68.1	Galectin-8	O00214	rs77189927	4	96495802	T	C	0.95	0.95	trans	-	-
LGI3.8003.57.3	Leucine-rich repeat LGI family member 3	Q8N145	rs10014926	4	45014454	T	A	0.07	0.07	trans	-	-
LHB.8376.25.4	Lutropin subunit beta	P01229	rs79609266	5	149527220	G	T	0.94	0.95	trans	-	-
LILRA4.8299.66.3	Leukocyte immunoglobulin-like receptor subfamily A member 4	P59901	rs78529238	5	106792201	T	C	0.97	0.97	trans	-	-
LILRA4.8299.66.3	Leukocyte immunoglobulin-like receptor subfamily A member 4	P59901	rs73283849	8	69705825	C	T	0.92	0.93	trans	-	-
LILRB2.5091.28.3	Leukocyte immunoglobulin-like receptor subfamily B member 2	Q8N423	rs4469949	12	6545611	G	A	0.65	0.65	trans	no	-
LILRB4.6453.70.3	Leukocyte immunoglobulin-like receptor subfamily B member 4	Q8NHJ6	rs2834904	21	36689996	C	T	0.60	0.60	trans	-	-
LINGO1.6620.82.3	Leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting protein 1	Q96FE5	rs35360713	3	120381896	G	A	0.88	0.88	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
LINGO3.10827.67.3	Leucine-rich repeat and immunoglobulin-like domain-containing nogo receptor-interacting protein 3	P0C6S8	rs17739723	5	4429269	G	A	0.74	0.75	trans	-	-
LONP1.6398.12.3	Lon protease homolog, mitochondrial	P36776	rs73226816	4	29476948	T	C	0.87	0.87	trans	-	-
LPCAT2.11175.45.3	Lysophosphatidylcholine acyltransferase 2	Q7L5N7	rs981138	14	27721363	C	A	0.29	0.29	trans	-	-
LRFN1.7910.41.3	Leucine-rich repeat and fibronectin type III domain-containing protein 1	Q9P244	rs139297576	4	146514810	A	T	0.98	0.97	trans	-	-
LRFN2.7200.4.3	Leucine-rich repeat and fibronectin type-III domain-containing protein 2	Q9ULH4	rs12890474	14	46480421	A	T	0.76	0.74	trans	-	-
LRIT2.11716.28.3	Leucine-rich repeat, immunoglobulin-like domain and transmembrane domain-containing protein 2	A6NDA9	rs6877907	5	180127017	T	A	0.04	0.05	trans	-	-
LRP1.10699.52.3	Prolow-density lipoprotein receptor-related protein 1	Q07954	rs4925433	22	49069015	T	C	0.89	0.89	trans	-	-
LRP1B.11275.94.3	Low-density lipoprotein receptor-related protein 1B	Q9NZR2	rs6675466	1	20338057	G	A	0.94	0.95	trans	-	no
LRPAP1.3640.14.3	alpha-2-macroglobulin receptor-associated protein	P30533	rs61880881	11	22270782	A	C	0.94	0.93	trans	-	no
LRPAP1.3640.14.3	alpha-2-macroglobulin receptor-associated protein	P30533	rs17438440	3	146579888	C	T	0.88	0.88	trans	no	-
LRPAP1.8829.4.3	alpha-2-macroglobulin receptor-associated protein	P30533	rs73161818	22	37573712	C	A	0.76	0.75	trans	-	no
LRRC3.6917.49.3	Leucine-rich repeat-containing protein 3	Q9BY71	rs72764980	16	6788326	C	T	0.97	0.97	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
LRRC37A2.8897.3.3	Leucine-rich repeat-containing protein 37A2	A6NM11	rs4143740	1	146697790	T	C	0.64	0.63	trans	-	-
LRRC4C.9369.174.3	Leucine-rich repeat-containing protein 4C	Q9HCJ2	rs114714046	3	114937671	A	G	0.98	0.98	trans	-	-
LRRN1.11293.14.3	Leucine-rich repeat neuronal protein 1	Q6UXK5	rs11996097	8	139152515	G	A	0.97	0.98	trans	no	-
LRRTM2.6904.14.3	Leucine-rich repeat transmembrane neuronal protein 2	O43300	rs3810459	20	61594353	A	G	0.46	0.46	trans	-	-
LRRTM4.6572.10.3	Leucine-rich repeat transmembrane neuronal protein 4	Q86VH4	rs45624432	10	16708368	C	A	0.95	0.96	trans	no	-
LRRTM4.8646.61.3	Leucine-rich repeat transmembrane neuronal protein 4	Q86VH4	rs4078399	2	71739391	A	G	0.82	0.82	trans	no	-
LRTM1.9368.64.3	Leucine-rich repeat and transmembrane domain-containing protein 1	Q9HBL6	rs35846682	6	41682486	C	T	0.90	0.91	trans	-	-
LTA.4703.87.2	Lymphotoxin-alpha	P01374	rs7841	2	135211861	C	T	0.90	0.89	trans	-	-
LY9.3324.51.1	T-lymphocyte surface antigen Ly-9	Q9HBG7	rs55993474	1	79557392	A	T	0.95	0.94	trans	-	-
LYSMD4.9106.87.3	LysM and putative peptidoglycan-binding domain-containing protein 4	Q5XG99	rs6973789	7	8956487	G	C	0.94	0.94	trans	-	-
MAGEA10.13610.9.3	Melanoma-associated antigen 10	P43363	rs17244095	9	19941355	C	A	0.84	0.83	trans	-	-
MAGEB10.11456.2.3	Melanoma-associated antigen B10	Q96LZ2	rs2365650	1	201311165	T	C	0.90	0.90	trans	-	-
MAGEB10.11456.2.3	Melanoma-associated antigen B10	Q96LZ2	rs13096279	3	132374053	G	A	0.07	0.06	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
MAN1A2.9077.10.3	Mannosyl-oligosaccharide 1,2-alpha-mannosidase IB	O60476	rs10407041	19	52342669	T	G	0.69	0.69	trans	-	-
MAP2K4.5242.37.3	Dual specificity mitogen-activated protein kinase kinase 4	P45985	rs6833765	4	19709	C	T	0.86	0.86	trans	-	-
MAP3K3.12990.39.3	Mitogen-activated protein kinase kinase kinase 3	Q99759	rs56292770	21	46810010	G	A	0.71	0.72	trans	-	-
MEGF10.11168.3.3	Multiple epidermal growth factor-like domains protein 10	Q96KG7	rs273701	18	23181382	C	A	0.52	0.52	trans	-	-
MEPE.3209.69.1	Matrix extracellular phosphoglycoprotein	Q9NQ76	rs6467467	7	132835532	G	A	0.17	0.16	trans	-	-
MET.2837.3.2	Hepatocyte growth factor receptor	P08581	rs4641389	10	97169467	C	T	0.50	0.50	trans	-	-
METTL1.12514.16.3	tRNA (guanine-N(7)-)-methyltransferase	Q9UBP6	rs35157786	8	106272609	A	G	0.72	0.71	trans	-	-
METTL1.12514.16.3	tRNA (guanine-N(7)-)-methyltransferase	Q9UBP6	rs34875393	8	52512866	G	C	0.83	0.85	trans	-	-
MFAP1.5606.24.3	Microfibrillar-associated protein 1	P55081	rs4770279	13	23041880	A	C	0.34	0.34	trans	-	-
MFAP1.5606.24.3	Microfibrillar-associated protein 1	P55081	rs7627256	3	11714521	C	A	0.43	0.44	trans	-	-
MFNG.5605.77.3	Beta-1,3-N-acetylglucosaminyltransferase manic fringe	O00587	rs28538862	4	188653234	G	A	0.25	0.24	trans	-	-
MGAT4C.7208.60.3	Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase C	Q9UBM8	rs17224975	11	97920349	T	C	0.94	0.93	trans	-	-
MILR1.11173.29.3	Allergin-1	Q7Z6M3	rs35648408	7	87391161	G	C	0.76	0.76	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
MLL2.13623.4.3	Histone-lysine N-methyltransferase 2D	O14686	rs62049594	16	77246517	C	A	0.58	0.58	trans	-	-
MMP10.3743.1.2	Stromelysin-2	P09238	rs13052751	21	44284413	C	T	0.58	0.56	trans	-	no
MMP16.9719.145.3	Matrix metalloproteinase-16	P51512	rs146137212	11	65501061	A	T	0.87	0.86	trans	-	no
MNX1.9576.58.3	Motor neuron and pancreas homeobox protein 1	P50219	rs1967318	5	98066468	G	A	0.49	0.49	trans	-	-
MPG.12438.127.3	DNA-3-methyladenine glycosylase	P29372	rs13021736	2	202115791	C	T	0.66	0.68	trans	-	-
MPZ.10615.18.3	Myelin protein P0	P25189	rs12692367	2	8763423	C	A	0.28	0.29	trans	-	-
MRAP.7895.108.3	Melanocortin-2 receptor accessory protein	Q8TCY5	rs1541837	8	16169243	A	C	0.63	0.63	trans	-	-
MRC1.2637.77.2	Macrophage mannose receptor 1	P22897	rs6578065	8	141073802	T	C	0.09	0.08	trans	-	-
MRPL52.7123.25.3	39S ribosomal protein L52, mitochondrial	Q86TS9	rs2292477	6	13346867	C	T	0.92	0.92	trans	-	-
MSLN.3893.64.1	Mesothelin	Q13421	rs115135029	3	13687922	T	A	0.94	0.94	trans	-	no
MSMB.10620.21.3	Beta-microseminoprotein	P08118	rs117036042	9	97612909	C	T	0.95	0.96	trans	-	-
MTMR1.11167.6.3	Myotubularin-related protein 1	Q13613	rs61906909	11	107762576	G	C	0.78	0.77	trans	-	-
MTMR1.11167.6.3	Myotubularin-related protein 1	Q13613	rs17022806	1	214574317	T	A	0.93	0.93	trans	-	-
MYCT1.13541.1.3	Myc target protein 1	Q8N699	rs10079672	5	148696383	G	A	0.82	0.84	trans	-	-
MYRF.8843.34.3	Myelin regulatory factor	Q9Y2G1	rs10478218	5	99260992	A	C	0.92	0.91	trans	-	-
MYZAP.9055.81.3	Myocardial zonula adherens protein	P0CAP1	rs7777594	7	29637619	G	A	0.96	0.96	trans	-	no
N6AMT1.11096.57.3	HemK methyltransferase family member 2	Q9Y5N5	rs80334360	7	25675528	A	G	0.88	0.89	trans	-	-
NAB2.13680.3.3	NGFI-A-binding protein 2	Q15742	rs2741088	8	6687235	G	A	0.81	0.80	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
NAPB.12655.30.3	Beta-soluble NSF attachment protein	Q9H115	rs146477950	21	32993934	T	G	0.96	0.96	trans	-	-
NCAM2.6507.16.3	Neural cell adhesion molecule 2	O15394	rs577380	3	13306257	T	C	0.33	0.33	trans	-	-
NCF2.10047.12.3	Neutrophil cytosol factor 2	P19878	rs143589027	1	224986444	C	T	0.89	0.90	trans	-	-
NCOA2.14045.12.3	Nuclear receptor coactivator 2	Q15596	rs995811	13	94008800	C	T	0.71	0.71	trans	-	-
NCR1.8360.169.3	Natural cytotoxicity triggering receptor 1	O76036	rs4696933	4	7985392	C	G	0.80	0.79	trans	-	no
NDRG4.7934.11.3	Protein NDRG4	Q9ULP0	rs187500955	11	63608204	C	T	0.97	0.97	trans	-	-
NDST1.6927.7.3	Bifunctional heparan sulfate N-deacetylase/N-sulfotransferase 1	P52848	rs10870218	10	134793832	C	T	0.43	0.42	trans	-	-
NDUFB4.10677.9.3	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 4	O95168	rs11791803	9	21156653	T	C	0.60	0.59	trans	-	-
NEFH.9900.36.3	Neurofilament heavy polypeptide	P12036	rs11256342	10	6057231	T	G	0.66	0.69	trans	-	-
NELL1.6544.33.3	Protein kinase C-binding protein NELL1	Q92832	rs2973926	5	72631179	A	C	0.83	0.83	trans	-	-
NET1.14260.112.3	Neuroepithelial cell-transforming gene 1 protein	Q7Z628	rs419301	2	32497236	C	G	0.56	0.55	trans	-	-
NGF.5801.72.3	beta-nerve growth factor	P01138	rs6686129	1	165591948	T	C	0.32	0.32	trans	-	-
NHP2L1.13602.6.3	NHP2-like protein 1	P55769	rs56292770	21	46810010	G	A	0.71	0.72	trans	-	-
NMB.9321.400.3	Neuromedin-B	P08949	rs2486962	1	203191920	G	A	0.52	0.51	trans	-	-
NME1.5909.51.3	Nucleoside diphosphate kinase A	P15531	rs7306771	12	104460934	T	G	0.64	0.64	trans	-	-
NOG.5846.24.3	Noggin	Q13253	rs9367080	6	40777596	A	T	0.65	0.66	trans	-	-
NOTUM.8252.2.3	Palmitoleoyl-protein carboxylesterase NOTUM	Q6P988	rs76989619	16	79711720	C	T	0.97	0.97	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
NOVA1.12557.18.3	RNA-binding protein Nova-1	P51513	rs338466	1	109486196	A	G	0.47	0.47	trans	-	-
NPPA.5443.62.2	Atrial natriuretic factor	P01160	rs6789159	3	2211932	A	G	0.32	0.32	trans	-	-
NPTX2.6521.35.3	Neuronal pentraxin-2	P47972	rs1859729	7	14296450	T	G	0.41	0.42	trans	-	-
NPW.9986.14.3	Neuropeptide W	Q8N729	rs2730927	12	38943171	C	T	0.95	0.94	trans	-	-
NRG1.9178.30.3	Neuregulin-1	Q02297	rs149839436	5	62567645	G	A	0.97	0.97	trans	-	-
NRG2.8060.7.3	Pro-neuregulin-2, membrane-bound isoform	O14511	rs4930806	12	5630381	A	G	0.09	0.09	trans	-	-
NRG4.14139.16.3	Neuregulin-4	Q8WWG1	rs1930594	9	17197991	C	G	0.81	0.81	trans	-	-
NRP1.3214.3.2	Neuropilin-1	O14786	rs7068721	10	112565606	C	T	0.61	0.60	trans	-	-
NUDT16L1.12497.29.3	Protein syndesmos	Q9BRJ7	rs2277413	12	9317784	A	G	0.70	0.71	trans	-	-
NUDT16L1.12497.29.3	Protein syndesmos	Q9BRJ7	rs10843767	12	9455501	T	C	0.59	0.61	trans	-	-
OBP2B.5680.54.3	Odorant-binding protein 2b	Q9NPH6	rs2127855	4	34495039	C	T	0.92	0.93	trans	-	-
OBP2B.5680.54.3	Odorant-binding protein 2b	Q9NPH6	rs12196876	6	73297052	G	A	0.85	0.85	trans	-	-
OCIAD1.3635.76.4	OCIA domain-containing protein 1	Q9NX40	rs74396411	4	181284516	G	A	0.92	0.93	trans	-	-
OLFML3.8660.5.3	Olfactomedin-like protein 3	Q9NRN5	rs16906040	9	2420424	C	T	0.96	0.96	trans	-	no
ORC6L.12389.4.3	Origin recognition complex subunit 6	Q9Y5N6	rs17116373	14	46176985	A	G	0.95	0.95	trans	-	-
ORC6L.12389.4.3	Origin recognition complex subunit 6	Q9Y5N6	rs11697735	20	3709225	G	A	0.69	0.69	trans	-	-
OSM.14063.17.3	Oncostatin-M	P13725	rs9360628	6	63335725	G	A	0.69	0.68	trans	no	-
P2RX6.7233.73.3	P2X purinoceptor 6	O15547	rs4872232	8	24279067	C	A	0.43	0.42	trans	-	-
PAIP1.12430.78.3	Polyadenylate-binding protein-interacting protein 1	Q9H074	rs10962917	9	17230086	G	A	0.83	0.83	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
PAK3.3387.1.2	Serine/threonine-protein kinase PAK 3	O75914	rs11047985	12	25529645	C	T	0.50	0.48	trans	-	-
PAPSS1.14007.22.3	Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 1	O43252	rs62271062	3	115521927	A	G	0.97	0.97	trans	-	-
PARP11.12882.7.3	Poly [ADP-ribose] polymerase 11	Q9NR21	rs3759481	13	42845556	G	C	0.89	0.90	trans	-	-
PATE4.8065.245.3	Prostate and testis expressed protein 4	P0C8F1	rs149437310	11	118226120	A	G	0.94	0.94	trans	-	-
PATE4.8065.245.3	Prostate and testis expressed protein 4	P0C8F1	rs875500	11	125698654	T	A	0.29	0.27	cis	-	-
PCBD2.6899.37.3	Pterin-4-alpha-carbinolamine dehydratase 2	Q9H0N5	rs74849342	9	36933549	G	T	0.95	0.95	trans	-	-
PCDHB1.9941.70.3	Protocadherin beta-1	Q9Y5F3	rs4800378	18	19835476	T	C	0.71	0.70	trans	-	-
PCDHGC5.7983.1.3	Protocadherin gamma-C5	Q9Y5F6	rs12206986	6	147444430	A	G	0.96	0.96	trans	-	-
PCOLCE2.6081.52.3	Procollagen C-endopeptidase enhancer 2	Q9UKZ9	rs543258852	5	58490298	C	A	0.97	0.97	trans	-	-
PCSK1.13388.57.3	Neuroendocrine convertase 1	P29120	rs9322474	6	154991447	C	T	0.43	0.42	trans	-	-
PDE6D.13491.40.3	Retinal rod rhodopsin-sensitive cGMP 3',5'-cyclic phosphodiesterase subunit delta	O43924	rs678793	3	159700115	A	G	0.89	0.88	trans	-	-
PDE9A.5201.50.4	High affinity cGMP-specific 3',5'-cyclic phosphodiesterase 9A	O76083	rs144606797	2	141508750	C	T	0.94	0.94	trans	-	-
PDGFRB.3459.49.2	Platelet-derived growth factor receptor beta	P09619	rs6442329	3	12817480	G	A	0.88	0.89	trans	-	-
PDIA3.7832.181.3	Protein disulfide-isomerase A3	P30101	rs10486426	7	21403759	A	C	0.92	0.93	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
PENK.9076.25.3	Proenkephalin-A	P01210	rs7650602	3	141147414	T	C	0.56	0.54	trans	-	-
PES1.4267.81.3	Pescadillo homolog	O00541	rs17045246	2	116667461	G	C	0.97	0.96	trans	-	-
PGM1.9173.21.3	Phosphoglucomutase-1	P36871	rs8129637	21	38151891	T	C	0.35	0.35	trans	-	-
PGRMC1.7863.50.3	Membrane-associated progesterone receptor component 1	O00264	rs6731777	2	168257448	A	T	0.92	0.92	trans	-	-
PHF3.11544.39.3	PHD finger protein 3	Q92576	rs17692625	2	212158902	A	G	0.90	0.90	trans	-	-
PIM1.5359.65.3	Serine/threonine-protein kinase pim-1	P11309	rs73121347	12	65831432	C	T	0.92	0.92	trans	-	-
PIN4.12718.43.3	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 4	Q9Y237	rs72829969	2	103476817	C	G	0.97	0.96	trans	-	-
PIWIL1.12793.4.3	Piwi-like protein 1	Q96J94	rs79537600	4	12069299	T	A	0.97	0.97	trans	-	-
PKLR.11203.97.3	Pyruvate kinase PKLR	P30613	rs7170123	15	61931318	T	G	0.08	0.08	trans	-	-
PLA2G2E.13670.81.3	Group IIE secretory phospholipase A2	Q9NZK7	rs6944897	7	68355951	T	A	0.96	0.96	trans	-	-
PLAU.4158.54.2	Urokinase-type plasminogen activator	P00749	rs6590199	11	126235667	G	A	0.83	0.84	trans	-	-
PMEL.6472.40.3	Melanocyte protein PMEL	P40967	rs541786053	2	44093231	G	A	0.96	0.97	trans	-	-
PMS2.11312.40.3	Mismatch repair endonuclease PMS2	P54278	rs62319809	4	129582846	T	A	0.67	0.70	trans	-	-
POGLUT1.6467.65.3	Protein O-glucosyltransferase 1	Q8NBL1	rs10911507	1	184071362	A	G	0.33	0.32	trans	-	-
POMC.4890.10.1	Corticotropin	P01189	rs10266911	7	147715709	T	C	0.68	0.67	trans	-	-
POR.2731.29.2	NADPH--cytochrome P450 reductase	P16435	rs78721608	7	22218403	A	C	0.96	0.96	trans	-	-
POSTN.6645.53.3	Periostin	Q15063	rs7586684	2	238791066	G	C	0.92	0.93	trans	no	-
PPCDC.13996.16.3	Phosphopantothenoylcysteine decarboxylase	Q96CD2	rs13014084	2	102854765	A	G	0.95	0.94	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
PPP3R1.3657.74.3	Calcineurin subunit B type 1	P63098	rs2898791	1	4334601	C	A	0.37	0.35	trans	-	-
PRDM1.14197.2.3	PR domain zinc finger protein 1	O75626	rs79916008	4	153777708	G	A	0.95	0.96	trans	-	-
PRIM1.13591.31.3	DNA primase small subunit	P49642	rs17733351	5	5603771	G	A	0.72	0.72	trans	-	-
PRNP.6545.58.3	Major prion protein	P04156	rs793715	6	24445131	T	C	0.04	0.05	trans	-	-
PROK1.2247.20.11	Prokineticin-1	P58294	rs79097477	6	168561725	G	A	0.94	0.94	trans	-	-
PRPSAP1.9478.69.3	Phosphoribosyl pyrophosphate synthase-associated protein 1	Q14558	rs112291783	13	101227999	C	T	0.95	0.94	trans	-	-
PRSS1.3049.61.2	Trypsin-1	P07477	rs7758999	6	110903974	A	G	0.44	0.42	trans	-	-
PSAPL1.8814.33.3	Proactivator polypeptide-like 1	Q6NUJ1	rs13223283	7	5226309	G	A	0.69	0.69	trans	-	-
PSG6.6456.17.3	Pregnancy-specific beta-1-glycoprotein 6	Q00889	rs201911713	2	209041644	T	C	0.97	0.97	trans	-	-
PSG8.6238.55.3	Pregnancy-specific beta-1-glycoprotein 8	Q9UQ74	rs2319921	8	97749482	C	A	0.22	0.22	trans	-	-
PSG9.9335.28.3	Pregnancy-specific beta-1-glycoprotein 9	Q00887	rs370045443	5	1228240	T	C	0.91	0.91	trans	-	-
PTH1R.13470.43.3	Parathyroid hormone/parathyroid hormone-related peptide receptor	Q03431	rs17428027	8	21709512	A	G	0.65	0.67	trans	-	-
PTH2.7257.18.3	Tuberoinfundibular peptide of 39 residues	Q96A98	rs11073564	15	86354674	C	A	0.32	0.32	trans	-	-
PTH2.7257.18.3	Tuberoinfundibular peptide of 39 residues	Q96A98	rs71446071	2	32707476	T	G	0.96	0.96	trans	-	-
PTK2B.8918.64.3	Protein-tyrosine kinase 2-beta	Q14289	rs75212570	14	72392830	C	A	0.89	0.88	trans	-	-
PTPRD.9296.15.3	Receptor-type tyrosine-protein phosphatase delta	P23468	rs11788900	9	89444841	A	G	0.96	0.96	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
PTS.12014.19.3	6-pyruvoyl tetrahydrobiopterin synthase	Q03393	rs139331832	1	71371112	A	G	0.96	0.96	trans	-	-
PTS.12014.19.3	6-pyruvoyl tetrahydrobiopterin synthase	Q03393	rs17606027	2	44103237	T	C	0.89	0.89	trans	-	-
PVRL2.6245.4.3	Nectin-2	Q92692	rs10145554	14	62234159	A	G	0.72	0.71	trans	-	-
PVRL3.13557.3.3	Nectin-3	Q9NQS3	rs560049307	20	11614995	A	G	0.92	0.93	trans	-	-
PYGL.11441.11.3	Glycogen phosphorylase, liver form	P06737	rs10909972	1	3492090	C	G	0.91	0.92	trans	-	-
QPCTL.8866.53.3	Glutaminyl-peptide cyclotransferase-like protein	Q9NXS2	rs7965184	12	9203359	G	C	0.63	0.65	trans	-	-
QPCTL.8866.53.3	Glutaminyl-peptide cyclotransferase-like protein	Q9NXS2	rs75931502	12	9436287	C	G	0.54	0.57	trans	-	-
QRFP.6463.59.3	Orexigenic neuropeptide QRFP	P83859	rs1879159	3	175752690	C	T	0.25	0.24	trans	-	-
QRFP.6463.59.3	Orexigenic neuropeptide QRFP	P83859	rs116116091	4	181360325	G	A	0.82	0.84	trans	-	-
RAB14.14283.12.3	Ras-related protein Rab-14	P61106	rs6864389	5	116481463	T	C	0.72	0.73	trans	-	-
RAB27B.13596.3.3	Ras-related protein Rab-27B	O00194	rs4147531	4	100212197	G	A	0.53	0.51	trans	-	-
RAD51L3.12554.10.3	DNA repair protein RAD51 homolog 4	O75771	rs28820590	4	69468634	C	T	0.37	0.35	trans	-	-
RAP2A.9885.41.3	Ras-related protein Rap-2a	P10114	rs17247004	4	97934806	A	G	0.79	0.79	trans	-	-
RAP2A.9885.41.3	Ras-related protein Rap-2a	P10114	rs185340883	9	139838949	C	A	0.97	0.96	trans	-	-
RBL1.12879.5.3	Retinoblastoma-like protein 1	P28749	rs2413588	22	39755175	A	G	0.30	0.32	trans	-	-
RBM19.11468.15.3	Probable RNA-binding protein 19	Q9Y4C8	rs77962070	1	60197393	G	A	0.96	0.96	trans	-	-
RBM24.11380.84.3	RNA-binding protein 24	Q9BX46	rs370225926	2	108538256	G	A	0.51	0.53	trans	-	-
RBM3.12747.89.3	RNA-binding protein 3	P98179	rs283454	5	153959567	A	G	0.54	0.53	trans	-	-
RBP7.14208.3.3	Retinoid-binding protein 7	Q96R05	rs9267845	6	32193698	A	T	0.63	0.63	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
REG3A.9277.16.3	Regenerating islet-derived protein 3-alpha	Q06141	rs34165183	7	38751445	C	T	0.91	0.90	trans	-	-
RELB.11464.9.3	Transcription factor RelB	Q01201	rs1935667	10	100890401	G	A	0.81	0.81	trans	-	-
RELL2.6600.70.3	RELT-like protein 2	Q8NC24	rs367621886	12	83801	T	A	0.41	0.40	trans	-	-
RELT.14112.40.3	Tumor necrosis factor receptor superfamily member 19L	Q969Z4	rs73482001	18	72997351	C	G	0.68	0.67	trans	-	no
REPIN1.13554.78.3	Replication initiator 1	Q9BWE0	rs188974339	3	8761751	T	C	0.94	0.94	trans	-	-
RET.3220.40.2	Proto-oncogene tyrosine-protein kinase receptor Ret	P07949	rs6853943	4	54403225	A	G	0.16	0.15	trans	-	-
RETN.3046.31.1	Resistin	Q9HD89	rs1357302	2	19758949	T	C	0.66	0.66	trans	-	-
RFESD.13603.7.3	Rieske domain-containing protein	Q8TAC1	rs144265151	20	52282382	T	G	0.88	0.88	trans	-	-
RFXAP.12726.3.3	Regulatory factor X-associated protein	O00287	rs111769779	14	22012746	G	A	0.93	0.92	trans	-	-
RGS10.11634.32.3	Regulator of G-protein signaling 10	O43665	rs112603185	4	169466549	A	G	0.96	0.96	trans	-	-
RHOD.12442.4.3	Rho-related GTP-binding protein RhoD	O00212	rs117842273	10	120358424	C	T	0.96	0.95	trans	-	-
RHOG.12540.25.3	Rho-related GTP-binding protein RhoG	P84095	rs6948789	7	103525524	C	T	0.79	0.78	trans	-	-
RIPK2.8993.151.3	Receptor-interacting serine/threonine-protein kinase 2	O43353	rs12668631	7	150495397	T	C	0.79	0.78	trans	no	-
RNASE2.8394.56.3	Non-secretory ribonuclease	P10153	rs4796505	17	6411883	C	G	0.89	0.88	trans	-	-
RNF148.7742.11.3	RING finger protein 148	Q8N7C7	rs200866866	5	40341519	A	C	0.85	0.84	trans	-	no
RNF43.14120.2.3	E3 ubiquitin-protein ligase RNF43	Q68DV7	rs201435186	2	241620918	C	G	0.87	0.87	trans	-	-
RNGTT.12847.27.3	mRNA-capping enzyme	O60942	rs74717621	11	94336503	G	A	0.97	0.96	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
RNPC3.12814.17.3	RNA-binding protein 40	Q96LT9	rs117174884	10	705114	C	T	0.96	0.95	trans	-	-
RNPC3.12814.17.3	RNA-binding protein 40	Q96LT9	rs78487989	12	6408626	G	A	0.90	0.90	trans	-	-
ROBO3.5117.14.3	Roundabout homolog 3	Q96MS0	rs11587874	1	233702138	T	C	0.46	0.45	trans	-	-
ROBO3.5117.14.3	Roundabout homolog 3	Q96MS0	rs316023	6	160681314	T	C	0.63	0.61	trans	-	-
ROCK2.13654.1.3	Rho-associated protein kinase 2	O75116	rs76279051	1	99804571	T	C	0.89	0.89	trans	-	-
ROR1.2590.69.4	Tyrosine-protein kinase transmembrane receptor ROR1	Q01973	rs6070721	20	57666181	T	C	0.88	0.87	trans	-	-
RPE.12646.2.3	Ribulose-phosphate 3-epimerase	Q96AT9	rs11640887	16	86051882	C	T	0.71	0.70	trans	-	-
RPL30.12478.15.3	60S ribosomal protein L30	P62888	rs56292770	21	46810010	G	A	0.71	0.72	trans	-	-
RPS7.3864.5.2	40S ribosomal protein S7	P62081	rs56292770	21	46810010	G	A	0.71	0.72	trans	-	-
RSPO3.13094.75.3	R-spondin-3	Q9BXY4	rs2054908	15	94069988	T	C	0.24	0.22	trans	-	-
RSPO4.8464.31.3	R-spondin-4	Q2I0M5	rs11014312	10	25229083	A	T	0.89	0.88	trans	-	-
RSPO4.8464.31.3	R-spondin-4	Q2I0M5	rs13325300	3	185999403	T	A	0.92	0.93	trans	-	-
RXFP1.14135.3.3	Relaxin receptor 1	Q9HBX9	rs4907192	1	84506658	T	C	0.95	0.95	trans	-	-
S100A13.7223.60.3	Protein S100-A13	Q99584	rs77919600	18	64709528	A	G	0.96	0.96	trans	-	-
S100A9.5339.49.3	Protein S100-A9	P06702	rs3132950	6	32194740	C	T	0.63	0.63	trans	-	-
SAMHD1.11303.7.3	Deoxynucleoside triphosphate triphosphohydrolase SAMHD1	Q9Y3Z3	rs13064030	3	127791949	A	G	0.96	0.96	trans	-	-
SAP18.12830.4.3	Histone deacetylase complex subunit SAP18	O00422	rs2110690	2	202185132	A	G	0.50	0.50	trans	-	-
SAP18.12830.4.3	Histone deacetylase complex subunit SAP18	O00422	rs2585132	8	143821088	G	C	0.83	0.82	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
SARS2.12348.46.3	Serine--tRNA ligase, mitochondrial	Q9NP81	rs62266384	3	116873590	C	T	0.52	0.54	trans	-	-
SAYSD1.6594.64.3	SAYSvFN domain-containing protein 1	Q9NPB0	rs79286296	17	73120670	C	G	0.96	0.96	trans	-	-
SCAMP5.13509.5.3	Secretory carrier-associated membrane protein 5	Q8TAC9	rs6467550	7	134378420	G	C	0.22	0.22	trans	-	-
SCARB2.5100.53.3	Lysosome membrane protein 2	Q14108	rs7578883	2	79074719	A	G	0.75	0.73	trans	-	-
SCARF2.5130.67.3	Scavenger receptor class F member 2	Q96GP6	rs200884603	16	78456786	A	T	0.93	0.93	trans	no	-
SCARF2.5130.67.3	Scavenger receptor class F member 2	Q96GP6	rs6074977	20	15931275	C	A	0.03	0.03	trans	no	-
SCARF2.9925.56.3	Scavenger receptor class F member 2	Q96GP6	rs2660665	8	137526799	G	T	0.59	0.57	trans	-	no
SCG3.7957.2.3	Secretogranin-3	Q8WXD2	rs7654912	4	82634029	T	C	0.17	0.16	trans	-	-
SCIN.12684.5.3	Adseverin	Q9Y6U3	rs6472278	8	67375492	A	G	0.91	0.92	trans	-	-
SCO1.7853.19.3	Protein SCO1 homolog, mitochondrial	O75880	rs6679253	1	60108811	A	T	0.16	0.15	trans	-	-
SEC13.14689.3.3	Protein SEC13 homolog	P55735	rs72794885	16	77915260	A	G	0.95	0.96	trans	-	-
SEC61G.8872.1.3	Protein transport protein Sec61 subunit gamma	P60059	rs1532330	5	43116752	T	C	0.29	0.30	trans	-	-
SELP.4154.57.2	P-selectin	P16109	rs77628528	3	13679906	G	A	0.96	0.95	trans	-	-
SEMA3G.5628.21.3	Semaphorin-3G	Q9NS98	rs7201599	16	89042367	C	T	0.61	0.60	trans	-	-
SEMA4D.5737.61.3	Semaphorin-4D	Q92854	rs4636031	6	18049822	A	G	0.82	0.83	trans	-	-
SERF1A.7842.52.3	Small EDRK-rich factor 1	O75920	rs10135970	14	98657339	A	G	0.96	0.96	trans	-	-
SERF1A.7842.52.3	Small EDRK-rich factor 1	O75920	rs4842885	15	86035194	A	G	0.06	0.06	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
SERPINA6.4785.30.3	Corticosteroid-binding globulin	P08185	rs9574691	13	81161654	C	T	0.59	0.58	trans	-	-
SERPINA6.4785.30.3	Corticosteroid-binding globulin	P08185	rs74396411	4	181284516	G	A	0.92	0.93	trans	-	-
SERPINF2.3024.18.2	Alpha-2-antiplasmin	P08697	rs62119636	19	8333648	G	C	0.84	0.83	trans	-	-
SERPING1.4479.14.2	Plasma protease C1 inhibitor	P05155	rs844138	8	104366138	G	A	0.44	0.43	trans	-	-
SFRP1.3221.54.1	Secreted frizzled-related protein 1	Q8N474	rs72711689	1	81707249	G	A	0.88	0.87	trans	-	-
SFRP1.3221.54.1	Secreted frizzled-related protein 1	Q8N474	rs12692019	2	140643312	A	T	0.80	0.79	trans	-	-
SFRP2.7751.121.3	Secreted frizzled-related protein 2	Q96HF1	rs3933269	6	39410851	T	C	0.93	0.93	trans	-	-
SH3BP2.7769.29.3	SH3 domain-binding protein 2	P78314	rs11088998	21	46810776	G	A	0.76	0.77	trans	-	-
SIRPB1.6247.9.3	Signal-regulatory protein beta-1	O00241	rs80329159	4	2435996	T	A	0.96	0.95	trans	-	-
SIRT5.12461.8.3	NAD-dependent protein deacylase sirtuin-5, mitochondrial	Q9NXA8	rs4238111	13	29103872	C	T	0.25	0.25	trans	-	-
SIT1.11194.6.3	Signaling threshold-regulating transmembrane adapter 1	Q9Y3P8	rs7787886	7	135589720	G	T	0.37	0.37	trans	-	-
SLAMF7.5487.7.3	SLAM family member 7	Q9NQ25	rs6956539	7	13960705	T	C	0.97	0.96	trans	no	-
SLC14A1.13430.50.3	Urea transporter 1	Q13336	rs3779034	7	148468617	C	T	0.47	0.47	trans	-	-
SLC35G2.13501.10.3	Solute carrier family 35 member G2	Q8TBE7	rs73052804	19	52155401	C	T	0.92	0.92	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
SLC5A8.13691.10.3	Sodium-coupled monocarboxylate transporter 1	Q8N695	rs28667269	4	108857386	A	G	0.66	0.66	trans	-	-
SLITRK4.7139.14.3	SLIT and NTRK-like protein 4	Q8IW52	rs75854564	11	123473031	G	A	0.54	0.53	trans	-	-
SLITRK4.7139.14.3	SLIT and NTRK-like protein 4	Q8IW52	rs73937193	18	5627994	A	G	0.85	0.83	trans	-	-
SMC3.14324.52.3	Structural maintenance of chromosomes protein 3	Q9UQE7	rs2023222	6	32194083	T	G	0.63	0.63	trans	-	-
SMCO2.7925.18.3	Single-pass membrane and coiled-coil domain-containing protein 2	A6NFE2	rs11210256	1	73931851	G	A	0.43	0.42	trans	-	-
SMIM9.8888.33.3	Small integral membrane protein 9	A6NGZ8	rs2515031	8	119490850	G	C	0.60	0.60	trans	-	-
SNAI2.10014.31.3	Zinc finger protein SNAI2	O43623	rs10795292	10	688424	T	C	0.11	0.10	trans	-	-
SNAI2.10014.31.3	Zinc finger protein SNAI2	O43623	rs75787785	12	15429855	C	G	0.97	0.96	trans	-	-
SNAI2.10014.31.3	Zinc finger protein SNAI2	O43623	rs3920702	9	71170130	G	A	0.45	0.46	trans	-	-
SNCA.8458.16.3	Alpha-synuclein	P37840	rs111330842	19	8203712	C	G	0.88	0.87	trans	-	-
SNRPA.12678.66.3	U1 small nuclear ribonucleoprotein A	P09012	rs11088998	21	46810776	G	A	0.76	0.77	trans	-	-
SOCS7.11657.86.3	Suppressor of cytokine signaling 7	O14512	rs5766023	22	45187309	C	T	0.51	0.51	trans	-	-
SORBS3.12976.49.3	Vinexin b	O60504	rs11038831	11	46266009	C	T	0.75	0.76	trans	-	-
SPACA3.8076.4.3	Sperm acrosome membrane-associated protein 3	Q8IXA5	rs137904288	5	144897002	G	A	0.95	0.95	trans	-	-
SPAG11B.5762.35.3	Sperm-associated antigen 11B	Q08648	rs1201976	20	59292496	A	G	0.44	0.44	trans	-	-
SPARC.3043.49.2	SPARC	P09486	rs9651726	11	61069358	T	C	0.88	0.88	trans	-	no
SPARC.3043.49.2	SPARC	P09486	rs116928246	12	125025742	G	A	0.54	0.55	trans	-	no

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
SPARCL1.13707.27.3	SPARC-like protein 1	Q14515	rs852109	20	17161173	C	G	0.12	0.12	trans	-	no
SPARCL1.4467.49.2	SPARC-like protein 1	Q14515	rs198805	6	26134672	G	A	0.15	0.15	trans	-	no
SPINT1.2828.82.2	Kunitz-type protease inhibitor 1	O43278	rs188628536	7	94595961	A	T	0.97	0.97	trans	-	-
SPINT3.7926.13.3	Kunitz-type protease inhibitor 3	P49223	rs6032259	20	44141331	A	G	0.47	0.44	cis	-	-
SPOCK3.9906.21.3	Testican-3	Q9BQ16	rs1874278	15	33959328	A	C	0.79	0.78	trans	-	-
SPOP.13727.44.3	Speckle-type POZ protein	O43791	rs67274280	10	56491559	G	A	0.89	0.90	trans	-	-
SPR.9287.6.3	Sepiapterin reductase	P35270	rs3819091	18	12364185	C	A	0.97	0.97	trans	-	-
SPSB1.13942.140.3	SPRY domain-containing SOCS box protein 1	Q96BD6	rs10981641	9	115900126	C	T	0.94	0.94	trans	-	-
SPTLC1.7886.26.3	Serine palmitoyltransferase 1	O15269	rs62336320	4	174630161	G	A	0.71	0.71	trans	-	-
SSRP1.5032.64.1	FACT complex subunit SSRP1	Q08945	rs6965978	7	49858495	A	C	0.73	0.74	trans	-	-
ST6GALNAC3.10626.116.3	Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 3	Q8NDV1	rs117220250	19	52430467	C	T	0.97	0.96	trans	no	-
STAB1.14599.18.3	Stabilin-1	Q9NY15	rs144793007	8	106828131	G	A	0.95	0.96	trans	-	-
STAU1.12471.47.3	Double-stranded RNA-binding protein Staufen homolog 1	O95793	rs113462565	2	170162904	C	T	0.89	0.89	trans	-	-
STC2.6231.46.3	Stanniocalcin-2	O76061	rs55970149	5	50923899	C	T	0.67	0.67	trans	-	-
STC2.6231.46.3	Stanniocalcin-2	O76061	rs10272049	7	24713542	C	T	0.97	0.98	trans	-	-
STK17B.5249.31.3	Serine/threonine-protein kinase 17B	O94768	rs6951678	7	1556765	C	T	0.52	0.53	trans	-	-
STMN2.10900.272.3	Stathmin-2	Q93045	rs9267845	6	32193698	A	T	0.63	0.63	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
STX6.10945.11.3	Syntaxin-6	O43752	rs12730100	1	100102285	A	G	0.87	0.86	trans	-	-
SULT2A1.9829.91.3	Bile salt sulfotransferase	Q06520	rs10472000	5	56514016	A	G	0.88	0.90	trans	-	-
SUMF1.6941.11.3	Sulfatase-modifying factor 1	Q8NBK3	rs115365541	4	157687170	C	A	0.97	0.97	trans	-	-
SUN5.11260.47.3	SUN domain-containing protein 5	Q8TC36	rs77806631	22	39595920	G	C	0.96	0.97	trans	-	-
SVEP1.11109.56.3	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	Q4LDE5	rs75366433	6	51127169	G	A	0.92	0.92	trans	no	-
SVEP1.11178.21.3	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	Q4LDE5	rs34159990	2	236436316	G	T	0.83	0.82	trans	-	no
SVEP1.11178.21.3	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	Q4LDE5	rs75274627	3	118901118	G	C	0.94	0.94	trans	-	no
SYT11.7089.42.3	Synaptotagmin-11	Q9BT88	rs1012334	21	34713751	A	T	0.53	0.54	trans	-	-
SYT11.7089.42.3	Synaptotagmin-11	Q9BT88	rs62336320	4	174630161	G	A	0.71	0.71	trans	-	-
SYT5.9099.19.3	Synaptotagmin-5	O00445	rs959065	9	105041206	A	G	0.45	0.44	trans	-	-
TACSTD2.14034.22.3	Tumor-associated calcium signal transducer 2	P09758	rs77189383	2	40657994	G	A	0.95	0.94	trans	no	-
TACSTD2.14034.22.3	Tumor-associated calcium signal transducer 2	P09758	rs17651293	5	75495279	G	A	0.80	0.81	trans	no	-
TACSTD2.2619.72.2	Tumor-associated calcium signal transducer 2	P09758	rs288613	16	62676177	A	G	0.39	0.37	trans	-	no
TAPBP.12378.71.3	Tapasin	O15533	rs56386479	13	95524146	C	T	0.92	0.91	trans	-	-
TAX1BP1.12721.4.3	Tax1-binding protein 1	Q86VP1	rs6886791	5	58411770	G	A	0.90	0.90	trans	-	-
TAX1BP3.12498.12.3	Tax1-binding protein 3	O14907	rs115530557	2	219915015	T	C	0.97	0.97	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
TEK.3773.15.4	Angiopoietin-1 receptor, soluble	Q02763	rs113995269	4	187812723	C	T	0.93	0.93	trans	-	-
TFF3.4721.54.2	Trefoil factor 3	Q07654	rs9260306	6	29916764	C	T	0.93	0.94	trans	-	yes
TFF3.4721.54.2	Trefoil factor 3	Q07654	rs2524277	6	31407579	G	A	0.94	0.94	trans	-	yes
TFF3.8323.163.3	Trefoil factor 3	Q07654	rs9260306	6	29916764	C	T	0.93	0.94	trans	-	yes
TFF3.8323.163.3	Trefoil factor 3	Q07654	rs2524277	6	31407579	G	A	0.94	0.94	trans	-	yes
TFRC.6895.1.3	Transferrin receptor protein 1	P02786	rs564548360	19	55682238	C	T	0.96	0.95	trans	no	-
TFRC.6895.1.3	Transferrin receptor protein 1	P02786	rs17662532	5	172508979	T	C	0.96	0.97	trans	no	-
THPO.5947.90.3	Thrombopoietin	P40225	rs75751545	19	55002628	A	G	0.87	0.87	trans	no	-
THSD7A.11174.8.3	Thrombospondin type-1 domain-containing protein 7A	Q9UPZ6	rs2359238	18	41278308	G	A	0.46	0.46	trans	-	-
TINAGL1.11192.168.3	Tubulointerstitial nephritis antigen-like	Q9GZM7	rs12445561	16	86609570	C	G	0.83	0.82	trans	-	-
TJP1.12001.7.3	Tight junction protein ZO-1	Q07157	rs6597545	9	133402652	G	C	0.70	0.70	trans	-	-
TMED10.6506.54.3	Transmembrane emp24 domain-containing protein 10	P49755	rs4589296	11	23780120	G	A	0.47	0.48	trans	-	-
TMEM132C.11128.29.3	Transmembrane protein 132C	Q8N3T6	rs2501941	9	78985430	C	T	0.14	0.13	trans	no	-
TMEM132C.7173.141.3	Transmembrane protein 132C	Q8N3T6	rs200508204	21	27251497	G	A	0.54	0.53	trans	-	no
TMEM132D.13416.8.3	Transmembrane protein 132D	Q14C87	rs8064039	16	49403920	T	C	0.66	0.67	trans	-	-
TMEM132D.13416.8.3	Transmembrane protein 132D	Q14C87	rs73187233	22	50611915	C	G	0.81	0.82	trans	-	-
TMEM230.11542.11.3	Transmembrane protein 230	Q96A57	rs6082968	20	2352590	C	T	0.33	0.32	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
TMEM230.11542.11.3	Transmembrane protein 230	Q96A57	rs56266868	7	150982495	G	A	0.96	0.95	trans	-	-
TMEM57.12367.52.3	Macoilin	Q8N5G2	rs11023770	11	15921656	C	G	0.71	0.73	trans	-	-
TMEM57.12367.52.3	Macoilin	Q8N5G2	rs12523973	6	53160233	T	A	0.92	0.94	trans	-	-
TMEM59L.9959.60.3	Transmembrane protein 59-like	Q9UK28	rs852079	20	17136774	T	C	0.25	0.24	trans	-	-
TMPRSS11B.10895.28.3	Transmembrane protease serine 11B	Q86T26	rs112352594	5	176168040	G	A	0.96	0.97	trans	-	-
TMPRSS11D.6547.83.3	Transmembrane protease serine 11D	O60235	rs12904548	15	87866035	G	T	0.09	0.08	trans	-	-
TMPRSS5.8002.27.3	Transmembrane protease serine 5	Q9H3S3	rs6037959	20	4741723	A	G	0.53	0.54	trans	-	-
TMPRSS5.8002.27.3	Transmembrane protease serine 5	Q9H3S3	rs140430972	7	27847984	C	T	0.97	0.97	trans	-	-
TNF.5936.53.3	Tumor necrosis factor	P01375	rs575184	11	96016375	G	A	0.73	0.71	trans	-	-
TNF.5936.53.3	Tumor necrosis factor	P01375	rs9265496	6	31297496	C	T	0.43	0.43	cis	-	-
TNFAIP6.5036.50.1	Tumor necrosis factor-inducible gene 6 protein	P98066	rs11771844	7	150675910	A	G	0.08	0.09	trans	-	-
TNFRSF10B.7693.13.3	Tumor necrosis factor receptor superfamily member 10B	O14763	rs12456737	18	36454112	G	A	0.37	0.37	trans	-	-
TNFRSF13B.2704.74.1	Tumor necrosis factor receptor superfamily member 13B	O14836	rs5829222	2	9796474	C	A	0.94	0.93	trans	-	-
TNFRSF1A.2654.19.1	Tumor necrosis factor receptor superfamily member 1A	P19438	rs12498735	4	184921299	C	T	0.55	0.55	trans	-	-
TNNI3.5441.67.3	Troponin I, cardiac muscle	P19429	rs74943542	3	42285235	A	G	0.64	0.64	trans	-	no

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
TNS2.11667.29.3	Tensin-2	Q63HR2	rs6912572	6	96715728	T	C	0.28	0.29	trans	-	-
TNS4.9927.96.3	Tensin-4	Q8IZW8	rs1548304	22	42691238	C	T	0.52	0.52	trans	-	-
TOPBP1.9947.22.3	DNA topoisomerase 2-binding protein 1	Q92547	rs6821	7	129774714	C	T	0.71	0.70	trans	-	-
TP53.6152.111.3	Cellular tumor antigen p53	P04637	rs2350913	2	6408208	G	A	0.24	0.24	trans	-	-
TPSG1.14103.12.3	Tryptase gamma	Q9NRR2	rs182162144	18	1477231	A	G	0.95	0.94	trans	-	-
TPST1.7928.183.3	Protein-tyrosine sulfotransferase 1	O60507	rs3759542	14	96133616	G	A	0.94	0.94	trans	-	-
TPST2.8024.64.3	Protein-tyrosine sulfotransferase 2	O60704	rs7734857	5	113110205	T	C	0.71	0.70	trans	-	-
TREML1.9329.28.3	Trem-like transcript 1 protein	Q86YW5	rs12509727	4	5470800	T	G	0.57	0.56	trans	-	-
TSN.12477.42.3	Translin	Q15631	rs2047422	1	55711967	A	G	0.36	0.38	trans	-	-
TSNARE1.7069.9.4	t-SNARE domain-containing protein 1	Q96NA8	rs4849274	2	114525193	G	A	0.67	0.68	trans	-	-
TSTA3.12657.2.3	GDP-L-fucose synthase	Q13630	rs62358068	5	8745671	A	C	0.71	0.70	trans	-	-
TXLNA.9065.28.3	Alpha-taxilin	P40222	rs77131492	2	187410496	A	G	0.96	0.96	trans	-	-
TXN.8813.160.3	Thioredoxin	P10599	rs274472	11	32676753	T	C	0.91	0.91	trans	-	-
TYK2.5260.80.3	Non-receptor tyrosine-protein kinase TYK2	P29597	rs73166407	22	37784229	G	T	0.81	0.80	trans	-	-
TYRO3.2611.72.2	Tyrosine-protein kinase receptor TYRO3	Q06418	rs62124843	2	17082794	C	T	0.97	0.97	trans	-	-
UBE2G2.9199.6.3	Ubiquitin-conjugating enzyme E2 G2	P60604	rs144562925	4	40278069	C	T	0.97	0.97	trans	-	-
UBE2I.2877.3.1	SUMO-conjugating enzyme UBC9	P63279	rs11088998	21	46810776	G	A	0.76	0.77	trans	-	-
UBXN4.9970.7.3	UBX domain-containing protein 4	Q92575	rs115910097	3	16954645	C	A	0.96	0.96	trans	-	no

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
UBXN4.9997.12.3	UBX domain-containing protein 4	Q92575	rs12693374	2	185206196	A	G	0.79	0.78	trans	no	-
UFM1.3836.51.2	Ubiquitin-fold modifier 1	P61960	rs11640887	16	86051882	C	T	0.71	0.70	trans	-	-
UGT1A8.8899.75.3	UDP-glucuronosyltransferase 1-8	Q9HAW9	rs9345317	6	93725600	T	C	0.77	0.76	trans	-	-
UNC5A.7975.97.3	Netrin receptor UNC5A	Q6ZN44	rs13088553	3	73626703	T	C	0.70	0.70	trans	-	-
UROS.11248.43.3	Uroporphyrinogen-III synthase	P10746	rs73134208	20	52289253	C	T	0.89	0.88	trans	-	-
USP25.9215.117.3	Ubiquitin carboxyl-terminal hydrolase 25	Q9UHP3	rs144858068	12	74241022	C	T	0.95	0.95	trans	-	-
VAMP3.7903.18.3	Vesicle-associated membrane protein 3	Q15836	rs7679436	4	3279745	C	T	0.78	0.78	trans	-	-
VCL.8750.46.3	Vinculin	P18206	rs10863126	12	86359857	T	C	0.67	0.67	trans	-	-
VTN.8280.238.3	Vitronectin	P04004	rs117128433	10	108299689	C	T	0.96	0.97	trans	no	-
VTN.8280.238.3	Vitronectin	P04004	rs9342196	6	90124635	C	T	0.80	0.79	trans	no	-
VWC2L.7995.16.3	von Willebrand factor C domain-containing protein 2-like	B2RUY7	rs12365739	11	14229813	C	T	0.82	0.83	trans	-	-
WFDC5.6969.14.3	WAP four-disulfide core domain protein 5	Q8TCV5	rs10089128	8	66651705	C	G	0.96	0.97	trans	-	-
WNK3.5493.17.3	Serine/threonine-protein kinase WNK3	Q9BYP7	rs56726300	2	171135616	C	T	0.94	0.93	trans	-	-
XDH.11264.33.3	Xanthine dehydrogenase/oxidase	P47989	rs113593110	1	211324535	C	T	0.92	0.92	trans	-	-
XRCC4.9886.28.3	DNA repair protein XRCC4	Q13426	rs10975139	9	5490027	C	T	0.61	0.61	trans	-	-

SOMAMER ID	Target full name	Uniprot	rsID	Chr	Pos	EA	NEA	Women EAF	Men EAF	Cis/ trans	Consistent other aptamers women	Consistent other aptamers men
YWHAB.YWHAE.YWHAG.YWHAH.YWHAQ.YWHAZ.SFN.4707.50.2	14-3-3 protein family	P27348,P31946,P31947,P61981,P62258,P63104,Q04917	rs10952071	7	7827076	T	C	0.86	0.88	trans	-	-
ZAP70.3837.6.2	Tyrosine-protein kinase ZAP-70	P43403	rs62402466	6	39696223	T	A	0.89	0.88	trans	no	-
ZFP91.13651.54.3	E3 ubiquitin-protein ligase ZFP91	Q96JP5	rs2505722	1	22542445	G	T	0.47	0.48	trans	-	-
ZNF174.11486.26.3	Zinc finger protein 174	Q15697	rs71357272	18	48643613	C	T	0.82	0.84	trans	-	-
ZNF23.11565.58.3	Zinc finger protein 23	P17027	rs1788484	21	44496941	C	T	0.68	0.68	trans	-	-
ZNF329.12803.9.3	Zinc finger protein 329	Q86UD4	rs7606391	2	183760854	A	C	0.61	0.61	trans	-	-
ZNF75D.11596.47.3	Zinc finger protein 75D	P51815	rs2193400	2	51400804	C	T	0.18	0.18	trans	-	-

Only pQTLs that were associated with the protein P_{GWAS}<1% FDR, passed the quality control (consistent effect directions P_{GWAS}<0.05 in both INTERVAL cohorts), and with $P_{sexdiff}$<1% FDR are included in the table due to space restrictions (Benjamini and Hochberg, 1995). EAFs are computed based on both cohorts. Cis/trans; a pQTL is considered to be cis if it is within 500 kb of any aptamer gene, otherwise it is considered trans. Consistent other aptamers denotes if a pQTL had consistent effect directions at P<1% FDR in the specified sex for all other aptamers targeting the same protein (Benjamini and Hochberg, 1995). "-" signifies that the analysis was not performed, either due to the pQTL not being an index SNP in that sex or because there were no other aptamers targeting the same protein or protein complex.

Appendix Table 3.2. Summary statistics for the pQTLs with sex-differential effects on proteins.

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
ACPL2.6079.59.3	rs7190621	T	Men	No	-0.041 (-0.128,0.046)	3.5E-01	0.243 (0.160,0.327)	1.0E-08	3.6E-06
ADAMTS3.8845.2.3	rs7277195	A	Women	No	0.220 (0.141,0.300)	5.8E-08	0.005 (-0.077,0.086)	9.1E-01	2.1E-04
ADAMTS4.2809.25.2	rs200641322	T	Men	No	0.057 (-0.100,0.213)	4.8E-01	-0.449 (-0.599,-0.299)	4.0E-09	4.6E-06
ADAMTS5.3168.8.2	rs4958481	G	Men	No	0.012 (-0.111,0.136)	8.5E-01	-0.359 (-0.486,-0.232)	3.1E-08	4.1E-05
ADCK4.9794.17.3	rs34962524	G	Men	No	0.043 (-0.049,0.136)	3.6E-01	-0.248 (-0.339,-0.157)	8.5E-08	1.0E-05
ADCYAP1.4593.11.1	rs7926118	G	Women	No	-0.231 (-0.313,-0.150)	2.4E-08	-0.007 (-0.086,0.072)	8.6E-01	1.1E-04
ADCYAP1.4593.11.1	rs11100189	G	Men	No	-0.004 (-0.079,0.070)	9.1E-01	0.197 (0.124,0.269)	9.1E-08	1.4E-04
ADCYAP1.8285.64.3	rs76713390	A	Men	No	-0.043 (-0.148,0.062)	4.2E-01	0.288 (0.184,0.393)	6.1E-08	1.1E-05
ADCYAP1.8285.64.3	rs61660593	G	Women	No	-0.477 (-0.649,-0.305)	5.6E-08	0.070 (-0.106,0.245)	4.4E-01	1.3E-05
ADGRF1.11243.90.3	rs6884348	C	Women	Yes	0.229 (0.152,0.306)	5.8E-09	-0.050 (-0.127,0.027)	2.0E-01	4.8E-07
ADGRF1.11243.90.3	rs55885888	G	Men	No	0.018 (-0.164,0.201)	8.5E-01	-0.500 (-0.681,-0.319)	6.4E-08	7.9E-05
ADGRF5.6409.57.3	rs10053843	A	Women	Yes	-0.192 (-0.262,-0.122)	8.3E-08	0.044 (-0.024,0.113)	2.0E-01	2.3E-06
ADH1B.9834.62.3	rs7808576	G	Men	No	-0.014 (-0.085,0.056)	6.9E-01	0.186 (0.119,0.254)	6.2E-08	5.7E-05
ADH4.8325.37.3	rs77672249	G	Men	Yes	-0.011 (-0.162,0.141)	8.9E-01	0.423 (0.269,0.577)	7.3E-08	8.2E-05
AGER.4125.52.2	rs79768452	A	Women	Yes	0.410 (0.260,0.561)	9.3E-08	-0.079 (-0.231,0.072)	3.0E-01	6.9E-06
AKR1B1.9854.36.3	rs11640887	C	Women	No	0.224 (0.147,0.302)	1.5E-08	-0.009 (-0.086,0.068)	8.2E-01	2.9E-05
ALDH3A1.11480.1.3	rs74418288	T	Men	No	-0.135 (-0.348,0.078)	2.1E-01	0.527 (0.337,0.718)	6.1E-08	5.6E-06
ALDH3B1.12940.35.3	rs56292770	G	Women	Yes	0.239 (0.160,0.318)	3.0E-09	-0.051 (-0.129,0.027)	2.0E-01	2.8E-07
ANAPC7.11690.47.3	rs3760003	C	Men	No	-0.030 (-0.100,0.039)	3.9E-01	0.189 (0.121,0.256)	4.2E-08	9.3E-06
ANGPT2.13660.76.3	rs73127904	G	Men	No	-0.001 (-0.158,0.155)	9.9E-01	0.407 (0.260,0.554)	5.3E-08	1.9E-04
ANGPT4.2500.2.3	rs35185467	C	Men	No	-0.122 (-0.201,-0.042)	2.8E-03	0.215 (0.138,0.292)	4.0E-08	2.6E-09
ANGPTL1.9092.33.3	rs55759365	A	Women	No	0.223 (0.146,0.300)	1.3E-08	-0.075 (-0.152,0.002)	5.6E-02	8.0E-08
ANKRD27.12445.50.3	rs535816341	T	Women	No	-0.631 (-0.860,-0.403)	5.9E-08	0.020 (-0.195,0.234)	8.6E-01	4.6E-05

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
ANKRD27.12445.50.3	rs12547239	G	Women	No	-0.234 (-0.311,-0.158)	2.1E-09	0.006 (-0.069,0.082)	8.7E-01	1.2E-05
ANKRD46.7851.30.3	rs75920135	G	Women	No	0.293 (0.188,0.397)	4.1E-08	-0.004 (-0.108,0.099)	9.3E-01	7.7E-05
ANP32B.4194.26.3	rs1651667	A	Women	No	-0.222 (-0.299,-0.145)	1.5E-08	0.009 (-0.066,0.083)	8.2E-01	2.4E-05
AP4M1.10076.1.3	rs76553855	T	Women	No	0.240 (0.155,0.326)	3.5E-08	0.017 (-0.067,0.101)	6.9E-01	2.6E-04
APBB2.12753.6.3	rs80029890	G	Women	No	-0.525 (-0.716,-0.334)	7.3E-08	-0.015 (-0.201,0.172)	8.8E-01	1.8E-04
APCDD1.6599.5.3	rs796500	A	Women	Yes	-0.398 (-0.541,-0.254)	5.7E-08	0.078 (-0.051,0.207)	2.3E-01	1.3E-06
APCS.2474.54.5	rs34827770	A	Men	No	0.044 (-0.073,0.160)	4.6E-01	-0.316 (-0.429,-0.203)	4.4E-08	1.4E-05
APLP2.10627.87.3	rs116928246	G	Men	No	0.037 (-0.034,0.107)	3.1E-01	-0.197 (-0.267,-0.127)	3.4E-08	4.1E-06
APOL1.11510.31.3	rs12193015	C	Women	No	0.224 (0.144,0.305)	4.8E-08	-0.014 (-0.092,0.065)	7.3E-01	3.4E-05
APOM.10445.20.3	rs17713196	C	Women	No	-0.364 (-0.450,-0.278)	1.6E-16	-0.118 (-0.202,-0.034)	6.1E-03	6.6E-05
APOM.14125.5.3	rs17581439	T	Women	No	0.216 (0.137,0.295)	7.5E-08	-0.020 (-0.096,0.055)	6.0E-01	2.2E-05
ARFIP1.13488.3.3	rs73047026	A	Men	No	0.036 (-0.048,0.119)	4.0E-01	-0.221 (-0.303,-0.140)	9.7E-08	1.5E-05
ARHGAP36.6289.78.3	rs12153655	C	Men	No	-0.046 (-0.133,0.041)	3.0E-01	0.235 (0.154,0.317)	1.6E-08	3.8E-06
ARHGAP36.6289.78.3	rs12339857	A	Women	No	-0.340 (-0.464,-0.216)	7.7E-08	-0.007 (-0.133,0.119)	9.1E-01	2.2E-04
ARHGEF1.13976.9.3	rs9364470	T	Women	No	-0.270 (-0.357,-0.182)	1.8E-09	-0.003 (-0.092,0.086)	9.5E-01	3.0E-05
ARID3A.3875.62.1	rs9537337	G	Women	Yes	0.397 (0.259,0.534)	1.7E-08	-0.135 (-0.277,0.007)	6.2E-02	1.4E-07
ARID3A.3875.62.1	rs1406742	T	Women	No	0.259 (0.169,0.349)	1.6E-08	-0.019 (-0.104,0.067)	6.7E-01	1.2E-05
ARRDC3.12352.70.3	rs13018449	C	Men	No	0.080 (-0.004,0.163)	6.1E-02	-0.225 (-0.306,-0.144)	4.8E-08	2.6E-07
ASGR1.5452.71.3	rs6431973	C	Women	No	0.198 (0.126,0.271)	8.8E-08	-0.025 (-0.097,0.048)	5.1E-01	2.2E-05
ASMTL.10089.7.3	rs73179258	G	Women	No	-0.633 (-0.856,-0.409)	2.9E-08	-0.067 (-0.259,0.124)	4.9E-01	1.7E-04
ATAD1.6625.31.3	rs61923795	G	Men	No	-0.042 (-0.172,0.088)	5.3E-01	0.343 (0.218,0.469)	8.2E-08	2.9E-05
ATAD1.6625.31.3	rs474122	A	Men	Yes	-0.092 (-0.172,-0.013)	2.2E-02	0.209 (0.132,0.286)	1.0E-07	8.5E-08
ATP4B.9994.217.3	rs117811535	C	Men	No	-0.013 (-0.224,0.197)	9.0E-01	0.523 (0.332,0.714)	8.3E-08	2.2E-04
ATP4B.9994.217.3	rs6502557	G	Women	No	-0.302 (-0.401,-0.203)	1.9E-09	-0.031 (-0.127,0.065)	5.3E-01	1.1E-04
AURKA.3091.70.2	rs79714874	A	Men	No	-0.166 (-0.348,0.016)	7.4E-02	0.528 (0.334,0.722)	9.4E-08	3.2E-07

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
AXIN2.12925.105.3	rs6559981	T	Women	No	0.228 (0.149,0.307)	1.6E-08	0.014 (-0.061,0.089)	7.2E-01	1.2E-04
AXIN2.8429.16.3	rs375029398	A	Women	No	0.505 (0.327,0.683)	2.9E-08	0.039 (-0.131,0.209)	6.5E-01	2.1E-04
AZGP1.9312.8.3	rs10077981	A	Men	Yes	-0.026 (-0.095,0.043)	4.5E-01	0.190 (0.123,0.256)	2.2E-08	1.0E-05
B3GNT8.9297.12.3	rs4895754	G	Women	Yes	-0.316 (-0.422,-0.210)	4.8E-09	0.071 (-0.031,0.173)	1.7E-01	2.3E-07
B4GALT6.10832.24.3	rs34932812	G	Women	Yes	-0.705 (-0.947,-0.463)	1.1E-08	0.045 (-0.199,0.288)	7.2E-01	1.9E-05
BACH1.12451.62.3	rs12642529	G	Men	No	-0.036 (-0.126,0.054)	4.3E-01	0.229 (0.145,0.312)	9.0E-08	2.5E-05
BAG3.10078.5.3	rs140735987	A	Men	No	0.053 (-0.101,0.207)	5.0E-01	-0.406 (-0.548,-0.264)	2.1E-08	1.8E-05
BCAN.3461.58.1	rs13023465	C	Men	No	0.026 (-0.065,0.117)	5.8E-01	0.261 (0.176,0.347)	2.0E-09	2.2E-04
BCAN.3461.58.1	rs1326906	T	Women	No	-0.728 (-0.988,-0.469)	3.7E-08	0.077 (-0.161,0.315)	5.3E-01	7.4E-06
BCL2.3412.7.1	rs1009850	A	Women	No	0.234 (0.156,0.312)	4.3E-09	-0.001 (-0.079,0.077)	9.8E-01	3.1E-05
BCL2A1.3413.50.2	rs13165690	G	Women	Yes	0.235 (0.150,0.319)	6.1E-08	-0.073 (-0.157,0.010)	8.5E-02	3.9E-07
BCL6.13640.5.3	rs9717986	C	Men	No	-0.058 (-0.181,0.065)	3.6E-01	0.327 (0.214,0.440)	1.4E-08	6.2E-06
BET1L.10959.125.3	rs80011000	A	Men	No	0.037 (-0.154,0.228)	7.0E-01	-0.480 (-0.654,-0.306)	6.0E-08	8.5E-05
BGLAP.11067.13.3	rs17187009	A	Women	No	-0.396 (-0.542,-0.250)	9.6E-08	0.084 (-0.056,0.224)	2.4E-01	3.3E-06
BGN.13690.26.3	rs13065384	C	Women	No	0.261 (0.170,0.351)	1.8E-08	0.012 (-0.076,0.100)	7.9E-01	1.2E-04
BMP10.3587.53.2	rs34998662	G	Men	No	0.004 (-0.167,0.175)	9.7E-01	0.499 (0.320,0.679)	5.1E-08	8.9E-05
BMP6.3736.60.3	rs6067937	C	Women	No	0.489 (0.310,0.668)	8.9E-08	-0.098 (-0.279,0.084)	2.9E-01	6.5E-06
BMPR1A.4859.6.1	rs11627959	A	Men	No	0.007 (-0.062,0.076)	8.4E-01	0.202 (0.133,0.270)	9.5E-09	9.1E-05
BNIP3L.7835.2.3	rs6840503	A	Women	No	-0.255 (-0.348,-0.162)	7.9E-08	0.024 (-0.068,0.116)	6.2E-01	3.0E-05
BRSK2.9790.28.3	rs661325	C	Men	No	-0.004 (-0.080,0.072)	9.2E-01	0.201 (0.128,0.274)	6.5E-08	1.4E-04
BTNL9.7950.142.3	rs71385947	G	Women	No	0.303 (0.193,0.413)	6.9E-08	-0.021 (-0.135,0.094)	7.3E-01	6.8E-05
C17orf78.6571.75.3	rs34164740	T	Men	No	0.072 (-0.010,0.154)	8.5E-02	-0.216 (-0.295,-0.138)	6.7E-08	6.5E-07
C17orf78.6571.75.3	rs6074977	C	Women	No	-0.546 (-0.741,-0.351)	4.3E-08	0.075 (-0.123,0.274)	4.6E-01	1.2E-05
C18orf32.8236.8.3	rs114441192	C	Women	No	-0.625 (-0.834,-0.417)	4.3E-09	-0.022 (-0.229,0.184)	8.3E-01	5.7E-05
C1orf198.8035.6.3	rs637914	T	Men	Yes	0.067 (-0.007,0.142)	7.7E-02	-0.204 (-0.277,-0.130)	5.4E-08	3.8E-07

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
C1orf226.7989.5.3	rs11498715	T	Men	No	0.023 (-0.077,0.124)	6.5E-01	-0.262 (-0.358,-0.166)	9.6E-08	6.0E-05
C6.4127.75.1	rs4237676	C	Men	No	0.003 (-0.066,0.073)	9.2E-01	0.187 (0.119,0.255)	7.2E-08	2.2E-04
CA9.3798.71.1	rs1234192	C	Men	Yes	0.026 (-0.046,0.098)	4.7E-01	-0.200 (-0.271,-0.129)	2.7E-08	1.1E-05
CACNA2D3.8885.6.3	rs12866815	T	Women	No	-0.227 (-0.303,-0.150)	5.9E-09	-0.017 (-0.094,0.060)	6.6E-01	1.4E-04
CACYBP.12432.23.3	rs12626681	G	Women	No	0.215 (0.136,0.294)	9.2E-08	-0.056 (-0.134,0.023)	1.6E-01	1.8E-06
CALB1.9918.23.3	rs1409177	A	Women	No	-0.538 (-0.731,-0.346)	4.2E-08	0.018 (-0.160,0.196)	8.4E-01	3.2E-05
CALN1.10933.107.3	rs9327896	T	Men	No	0.064 (-0.006,0.133)	7.5E-02	-0.189 (-0.256,-0.121)	4.3E-08	3.7E-07
CALR.5264.65.3	rs117799470	C	Women	No	0.566 (0.361,0.770)	6.0E-08	-0.129 (-0.350,0.093)	2.6E-01	6.4E-06
CALR.5264.65.3	rs957327	A	Men	No	-0.001 (-0.071,0.069)	9.7E-01	-0.206 (-0.276,-0.136)	7.5E-09	5.0E-05
CAPN2.14684.17.3	rs34685614	T	Women	No	-0.452 (-0.615,-0.289)	5.7E-08	0.091 (-0.071,0.253)	2.7E-01	3.6E-06
CARHSP1.12808.103.3	rs11640887	C	Women	No	0.227 (0.150,0.305)	9.6E-09	-0.033 (-0.110,0.045)	4.1E-01	3.2E-06
CARS.14098.28.3	rs553750511	T	Women	No	-0.650 (-0.872,-0.429)	9.0E-09	0.147 (-0.059,0.352)	1.6E-01	2.4E-07
CARS.14098.28.3	rs4561701	A	Men	No	0.042 (-0.038,0.122)	3.0E-01	-0.214 (-0.292,-0.135)	8.9E-08	7.5E-06
CASP10.5340.24.3	rs11835364	C	Men	Yes	0.095 (-0.054,0.244)	2.1E-01	-0.434 (-0.592,-0.277)	6.7E-08	1.7E-06
CASS4.12855.16.3	rs12694096	T	Women	No	0.317 (0.201,0.434)	9.2E-08	-0.031 (-0.149,0.088)	6.1E-01	4.0E-05
CAT.3488.64.2	rs62453446	T	Men	No	0.016 (-0.089,0.122)	7.6E-01	-0.295 (-0.400,-0.190)	3.7E-08	4.2E-05
CBL.12016.60.3	rs11088999	C	Women	Yes	0.220 (0.141,0.299)	4.9E-08	-0.017 (-0.095,0.061)	6.7E-01	2.8E-05
CBLN1.9313.27.3	rs17157815	T	Men	No	-0.020 (-0.188,0.147)	8.1E-01	-0.526 (-0.700,-0.353)	2.7E-09	3.9E-05
CBS.10086.39.3	rs7312302	A	Women	No	0.559 (0.363,0.756)	2.3E-08	-0.057 (-0.230,0.116)	5.2E-01	3.8E-06
CCBL2.12682.5.3	rs11195705	C	Women	No	0.535 (0.350,0.719)	1.3E-08	-0.141 (-0.332,0.051)	1.5E-01	6.4E-07
CCL1.13687.5.3	rs113241089	G	Women	No	0.570 (0.368,0.772)	3.1E-08	-0.057 (-0.232,0.119)	5.3E-01	4.4E-06
CCL1.13687.5.3	rs12424372	T	Women	No	0.231 (0.146,0.315)	9.5E-08	-0.043 (-0.128,0.042)	3.2E-01	7.6E-06
CCL19.4922.13.1	rs13188960	G	Women	No	0.211 (0.133,0.289)	9.6E-08	-0.010 (-0.086,0.065)	7.9E-01	6.0E-05
CCL21.2516.57.3	rs73085438	C	Women	No	0.327 (0.206,0.447)	1.1E-07	-0.013 (-0.142,0.116)	8.4E-01	1.6E-04
CD177.13116.25.3	rs17421832	C	Women	Yes	0.281 (0.178,0.384)	8.9E-08	-0.027 (-0.132,0.078)	6.2E-01	4.1E-05

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
CD226.5721.1.3	rs404449	T	Men	No	0.018 (-0.056,0.092)	6.3E-01	-0.197 (-0.268,-0.125)	6.4E-08	4.1E-05
CD300A.5630.48.3	rs3872610	C	Men	No	-0.017 (-0.107,0.074)	7.2E-01	0.242 (0.155,0.328)	5.0E-08	5.3E-05
CD300E.10798.4.3	rs6729654	A	Women	No	0.244 (0.167,0.321)	5.6E-10	0.015 (-0.060,0.090)	7.0E-01	3.0E-05
CD3G.7924.7.3	rs9571482	A	Women	No	-0.209 (-0.279,-0.138)	6.8E-09	-0.008 (-0.077,0.062)	8.2E-01	7.0E-05
CD4.3143.3.1	rs4811121	G	Men	No	0.037 (-0.038,0.112)	3.4E-01	-0.204 (-0.278,-0.130)	6.7E-08	7.9E-06
CD44.9283.8.3	rs12901497	A	Women	No	0.226 (0.144,0.307)	6.1E-08	-0.023 (-0.104,0.057)	5.7E-01	2.1E-05
CD5.14065.11.3	rs34169171	C	Men	No	0.031 (-0.156,0.219)	7.4E-01	-0.565 (-0.773,-0.357)	1.0E-07	3.0E-05
CD5.14065.11.3	rs10469849	T	Men	No	0.046 (-0.090,0.182)	5.1E-01	-0.368 (-0.503,-0.234)	8.4E-08	2.3E-05
CD58.10938.13.3	rs8096899	G	Women	No	0.452 (0.290,0.614)	4.4E-08	-0.157 (-0.323,0.008)	6.2E-02	2.4E-07
CD74.6974.6.3	rs4785468	T	Men	No	-0.076 (-0.146,-0.006)	3.3E-02	0.188 (0.121,0.256)	4.6E-08	9.9E-08
CD96.9735.44.3	rs893874	G	Men	Yes	0.059 (-0.048,0.165)	2.8E-01	-0.291 (-0.395,-0.187)	4.7E-08	4.3E-06
CD97.2822.56.2	rs7953479	G	Men	No	0.050 (-0.021,0.122)	1.7E-01	-0.184 (-0.250,-0.118)	4.6E-08	2.4E-06
CDH2.3797.1.1	rs1943603	T	Men	No	-0.083 (-0.179,0.013)	9.1E-02	0.251 (0.159,0.343)	9.5E-08	9.1E-07
CDH2.3797.1.1	rs28421485	G	Men	No	0.005 (-0.068,0.078)	8.9E-01	-0.204 (-0.276,-0.131)	3.8E-08	7.0E-05
CDKN1B.3719.2.2	rs4819094	T	Women	Yes	0.231 (0.150,0.312)	2.1E-08	-0.052 (-0.131,0.028)	2.0E-01	9.8E-07
CDON.4541.49.2	rs140593096	C	Women	No	-0.583 (-0.791,-0.374)	4.2E-08	0.065 (-0.133,0.262)	5.2E-01	9.7E-06
CEACAM21.7204.1.3	rs78074431	G	Men	No	-0.065 (-0.231,0.101)	4.4E-01	0.425 (0.280,0.570)	9.1E-09	1.3E-05
CELA2A.7140.1.3	rs74366845	C	Men	No	0.002 (-0.084,0.089)	9.6E-01	-0.240 (-0.325,-0.155)	3.2E-08	9.0E-05
CELA2A.7140.1.3	rs9831862	C	Men	No	-0.013 (-0.085,0.058)	7.2E-01	0.198 (0.130,0.266)	1.1E-08	2.7E-05
CFAP45.6553.68.3	rs10984054	T	Women	No	-0.345 (-0.468,-0.222)	4.0E-08	0.041 (-0.083,0.165)	5.1E-01	1.5E-05
CFHR5.3666.17.4	rs2474033	C	Women	No	-0.480 (-0.650,-0.310)	3.3E-08	-0.019 (-0.199,0.161)	8.4E-01	2.6E-04
CGA.FSHB.3032.11.2	rs6790495	C	Men	No	0.008 (-0.065,0.081)	8.3E-01	-0.222 (-0.295,-0.149)	2.5E-09	1.3E-05
CGA.LHB.2953.31.2	rs78537284	G	Women	No	0.655 (0.530,0.779)	7.7E-25	1.188 (1.071,1.305)	4.5E-88	9.7E-10
CGA.LHB.2953.31.2	rs139643250	C	Men	No	0.654 (0.528,0.779)	1.7E-24	1.212 (1.095,1.330)	8.1E-92	1.7E-10
CGA.LHB.2953.31.2	rs12530875	G	Men	No	0.032 (-0.114,0.178)	6.7E-01	-0.393 (-0.531,-0.255)	2.4E-08	3.5E-05

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
CGA.TSHB.3521.16.2	rs112721893	C	Men	No	0.055 (-0.016,0.126)	1.3E-01	-0.203 (-0.273,-0.133)	1.2E-08	4.0E-07
CHCHD10.11270.17.3	rs146728922	T	Men	Yes	-0.126 (-0.311,0.060)	1.8E-01	0.523 (0.334,0.711)	5.4E-08	1.5E-06
CHGB.8235.48.3	rs10788525	C	Women	No	-0.196 (-0.266,-0.125)	4.6E-08	0.023 (-0.046,0.092)	5.1E-01	1.3E-05
CHKB.13117.232.3	rs6725784	G	Men	No	0.159 (-0.032,0.350)	1.0E-01	-0.465 (-0.635,-0.295)	8.5E-08	1.8E-06
CHST15.14097.86.3	rs11252810	C	Men	No	-0.018 (-0.157,0.121)	8.0E-01	-0.413 (-0.556,-0.270)	1.4E-08	1.0E-04
CHST3.7189.55.3	rs58583902	T	Men	No	-0.049 (-0.225,0.126)	5.8E-01	0.457 (0.290,0.623)	7.3E-08	4.1E-05
CHST9.11646.4.3	rs1434887	T	Women	No	-0.326 (-0.435,-0.216)	5.0E-09	-0.015 (-0.132,0.102)	8.0E-01	1.4E-04
CLEC2A.10953.14.3	rs6876150	A	Men	No	0.050 (-0.018,0.119)	1.5E-01	-0.204 (-0.272,-0.136)	4.4E-09	2.6E-07
CLEC7A.3603.60.5	rs116862758	A	Men	No	0.001 (-0.208,0.210)	9.9E-01	0.540 (0.343,0.736)	7.4E-08	2.3E-04
CLEC7A.3603.60.5	rs11716867	C	Men	No	-0.015 (-0.092,0.062)	7.0E-01	0.202 (0.129,0.276)	6.8E-08	6.6E-05
CLIC5.12475.48.3	rs11981812	G	Men	Yes	-0.040 (-0.132,0.052)	3.9E-01	0.255 (0.167,0.342)	1.2E-08	5.2E-06
CLINT1.11659.31.3	rs113158475	C	Men	No	-0.056 (-0.199,0.088)	4.5E-01	0.408 (0.263,0.554)	4.0E-08	8.7E-06
CLMP.10440.26.3	rs911148	C	Women	No	-0.204 (-0.278,-0.130)	5.9E-08	0.031 (-0.042,0.105)	4.0E-01	9.1E-06
CLN5.9795.9.3	rs566185205	A	Women	No	-0.333 (-0.452,-0.214)	4.1E-08	0.017 (-0.099,0.132)	7.7E-01	3.5E-05
CLPS.5749.53.3	rs242057	G	Women	No	-0.229 (-0.307,-0.150)	1.1E-08	-0.006 (-0.084,0.073)	8.8E-01	8.2E-05
CMA1.3423.59.2	rs11580550	A	Women	No	-0.260 (-0.346,-0.173)	3.7E-09	-0.009 (-0.095,0.077)	8.4E-01	5.5E-05
CMPK1.8312.139.3	rs2248409	A	Women	No	0.195 (0.125,0.265)	4.5E-08	-0.009 (-0.076,0.059)	8.0E-01	3.8E-05
COA3.7888.58.3	rs9974464	C	Women	No	0.259 (0.165,0.352)	5.4E-08	-0.011 (-0.101,0.079)	8.1E-01	4.4E-05
COG8.9543.131.3	rs11216799	G	Women	Yes	0.335 (0.215,0.455)	4.2E-08	-0.079 (-0.195,0.037)	1.8E-01	1.1E-06
COL11A2.11278.4.3	rs148930599	C	Men	No	0.042 (-0.080,0.164)	5.0E-01	-0.325 (-0.443,-0.206)	8.5E-08	2.4E-05
COL1A1.11140.56.3	rs3123687	C	Women	No	0.419 (0.270,0.569)	3.6E-08	0.025 (-0.116,0.165)	7.3E-01	1.6E-04
COL1A1.11140.56.3	rs114160127	G	Women	No	0.684 (0.446,0.923)	1.9E-08	0.087 (-0.117,0.292)	4.0E-01	2.0E-04
COL20A1.8804.39.3	rs12442553	C	Men	No	-0.047 (-0.246,0.151)	6.4E-01	0.528 (0.337,0.720)	6.4E-08	4.3E-05
COL6A1.11150.3.3	rs150370694	C	Women	No	-0.495 (-0.675,-0.315)	6.9E-08	-0.007 (-0.191,0.177)	9.4E-01	2.1E-04
COL8A1.4807.13.3	rs113936403	G	Men	No	0.127 (-0.085,0.338)	2.4E-01	-0.557 (-0.757,-0.358)	4.1E-08	3.9E-06

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
COMMD7.2823.7.1	rs77341824	T	Men	No	0.078 (-0.031,0.186)	1.6E-01	-0.302 (-0.406,-0.199)	1.1E-08	6.7E-07
COTL1.4905.63.1	rs34592426	G	Women	Yes	0.302 (0.206,0.398)	7.6E-10	-0.039 (-0.130,0.052)	4.0E-01	4.4E-07
COX5B.7887.57.3	rs2884815	T	Women	No	-0.213 (-0.285,-0.141)	6.8E-09	0.015 (-0.055,0.085)	6.8E-01	8.9E-06
CPB1.6356.3.3	rs78040619	C	Men	No	-0.003 (-0.198,0.191)	9.7E-01	-0.502 (-0.685,-0.319)	7.9E-08	2.6E-04
CPVL.8891.7.3	rs2686200	A	Women	No	-0.238 (-0.314,-0.162)	7.7E-10	-0.014 (-0.089,0.061)	7.2E-01	4.0E-05
CPXM1.6255.74.3	rs116928246	G	Men	No	0.016 (-0.054,0.087)	6.5E-01	-0.190 (-0.260,-0.120)	9.7E-08	4.8E-05
CPZ.6493.9.3	rs11099939	G	Women	No	0.546 (0.350,0.743)	5.3E-08	0.006 (-0.191,0.202)	9.5E-01	1.4E-04
CROT.13929.27.3	rs36113849	T	Men	Yes	-0.119 (-0.253,0.015)	8.1E-02	0.388 (0.247,0.528)	6.7E-08	3.2E-07
CRYBB2.10000.28.3	rs2282160	T	Men	No	-0.049 (-0.120,0.023)	1.8E-01	0.194 (0.126,0.262)	2.3E-08	1.5E-06
CSPG4.8951.162.3	rs7929804	G	Women	No	0.202 (0.132,0.272)	1.7E-08	0.015 (-0.054,0.084)	6.6E-01	2.1E-04
CTNNA2.9872.23.3	rs10738148	C	Men	No	-0.050 (-0.123,0.022)	1.7E-01	0.204 (0.133,0.275)	1.9E-08	9.5E-07
CTSF.9212.22.3	rs36012476	G	Women	No	-0.485 (-0.650,-0.320)	7.8E-09	0.062 (-0.111,0.236)	4.8E-01	7.3E-06
CTSV.3364.76.2	rs62512308	G	Men	No	-0.017 (-0.191,0.158)	8.5E-01	0.457 (0.293,0.621)	4.7E-08	1.0E-04
CTSV.3364.76.2	rs77290088	G	Men	No	-0.007 (-0.215,0.201)	9.5E-01	-0.548 (-0.749,-0.347)	9.0E-08	2.4E-04
CUL3.10045.47.3	rs215372	A	Women	No	-0.206 (-0.282,-0.131)	8.4E-08	0.016 (-0.057,0.089)	6.7E-01	3.3E-05
CUL4B.13743.56.3	rs10993060	G	Men	No	0.018 (-0.077,0.114)	7.1E-01	-0.304 (-0.400,-0.209)	4.5E-10	2.9E-06
CXCL11.3038.9.2	rs35170645	A	Women	No	0.327 (0.258,0.395)	1.4E-20	0.134 (0.066,0.202)	1.1E-04	9.2E-05
CXCL12.2330.2.1	rs28604315	A	Women	No	-0.221 (-0.301,-0.140)	7.6E-08	0.019 (-0.064,0.102)	6.5E-01	4.7E-05
CXCL13.3487.32.2	rs117939745	C	Men	No	0.067 (-0.116,0.249)	4.7E-01	-0.493 (-0.669,-0.316)	4.6E-08	1.6E-05
CXCL16.2436.49.4	rs71586562	C	Men	Yes	0.023 (-0.057,0.103)	5.7E-01	-0.227 (-0.302,-0.151)	4.0E-09	8.8E-06
CXCL5.2979.8.2	rs9853250	C	Women	No	0.205 (0.131,0.279)	5.3E-08	-0.034 (-0.104,0.036)	3.4E-01	4.2E-06
CXCL9.11593.21.3	rs2470174	C	Men	No	-0.032 (-0.158,0.094)	6.2E-01	0.327 (0.207,0.447)	9.1E-08	5.1E-05
CYGB.11546.7.3	rs75067265	C	Men	No	0.028 (-0.162,0.219)	7.7E-01	-0.527 (-0.712,-0.341)	2.6E-08	4.2E-05
DCAF5.11283.13.3	rs1428250	G	Men	No	0.086 (-0.095,0.267)	3.5E-01	-0.497 (-0.677,-0.317)	5.8E-08	7.2E-06
DCN.2666.53.2	rs2515031	G	Men	Yes	0.051 (-0.020,0.121)	1.6E-01	-0.193 (-0.262,-0.123)	5.7E-08	1.6E-06

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
DCP1A.14008.22.3	rs72837699	C	Women	No	0.354 (0.224,0.484)	9.7E-08	-0.003 (-0.121,0.114)	9.5E-01	6.5E-05
DCUN1D3.13553.4.3	rs35269401	G	Women	No	0.279 (0.178,0.379)	5.8E-08	0.012 (-0.090,0.114)	8.1E-01	2.7E-04
DCUN1D3.13553.4.3	rs4524627	A	Women	No	0.305 (0.198,0.412)	2.0E-08	-0.074 (-0.182,0.033)	1.7E-01	8.5E-07
DDX39B.9742.59.3	rs56292770	G	Women	Yes	0.215 (0.136,0.294)	9.5E-08	-0.016 (-0.093,0.062)	6.9E-01	4.4E-05
DDX39B.9742.59.3	rs2023222	T	Men	No	0.001 (-0.072,0.073)	9.8E-01	0.198 (0.129,0.267)	2.1E-08	1.1E-04
DGCR6.9444.70.3	rs7024075	C	Women	Yes	0.216 (0.139,0.294)	4.0E-08	-0.041 (-0.117,0.035)	2.9E-01	3.2E-06
DHFR.9823.2.3	rs6910044	A	Women	Yes	-0.199 (-0.268,-0.129)	2.5E-08	0.020 (-0.051,0.090)	5.8E-01	1.6E-05
DHFR.9823.2.3	rs55957950	A	Women	No	-0.211 (-0.287,-0.134)	6.3E-08	0.033 (-0.045,0.111)	4.1E-01	1.2E-05
DIABLO.3122.6.2	rs10250463	G	Men	No	0.006 (-0.096,0.108)	9.1E-01	-0.286 (-0.384,-0.189)	9.7E-09	5.1E-05
DLL1.5349.69.3	rs9295988	C	Men	No	0.011 (-0.128,0.149)	8.8E-01	0.390 (0.255,0.525)	1.4E-08	1.2E-04
DNAJB9.11214.40.3	rs76855778	A	Women	No	0.486 (0.309,0.663)	7.8E-08	0.006 (-0.163,0.175)	9.4E-01	1.2E-04
DNAJC17.14655.1.3	rs58914460	C	Women	No	-0.232 (-0.317,-0.146)	1.0E-07	-0.000 (-0.085,0.085)	1.0e+00	1.7E-04
DNAJC18.8033.1.3	rs9841975	A	Women	No	-0.204 (-0.276,-0.133)	2.0E-08	0.009 (-0.063,0.081)	8.0E-01	3.6E-05
DNAJC4.8016.19.3	rs703473	G	Men	No	0.024 (-0.149,0.196)	7.9E-01	0.469 (0.307,0.630)	1.4E-08	2.3E-04
DNAJC4.8016.19.3	rs4858378	A	Women	No	-0.227 (-0.309,-0.144)	7.5E-08	0.031 (-0.050,0.112)	4.6E-01	1.3E-05
DNER.11442.1.3	rs7249288	T	Men	No	0.003 (-0.087,0.093)	9.4E-01	-0.257 (-0.346,-0.168)	1.7E-08	5.8E-05
DPEP1.8794.13.3	rs59423625	C	Women	No	0.393 (0.263,0.524)	3.4E-09	-0.023 (-0.149,0.104)	7.3E-01	7.3E-06
DPYSL3.12707.26.3	rs11646557	A	Women	No	0.210 (0.133,0.287)	8.3E-08	-0.013 (-0.087,0.062)	7.4E-01	4.8E-05
DRGX.8034.6.3	rs75678034	C	Women	No	0.362 (0.230,0.495)	8.0E-08	-0.022 (-0.154,0.110)	7.5E-01	5.5E-05
DSC2.13126.52.3	rs533379	G	Men	No	0.004 (-0.070,0.078)	9.1E-01	0.202 (0.129,0.275)	6.3E-08	1.9E-04
DSC3.4981.6.1	rs2428085	C	Women	No	-0.399 (-0.535,-0.264)	7.2E-09	0.049 (-0.085,0.182)	4.7E-01	3.7E-06
DSG1.2976.58.2	rs113823715	T	Men	Yes	-0.047 (-0.174,0.080)	4.7E-01	0.335 (0.216,0.454)	3.5E-08	1.7E-05
DUSP4.10035.6.3	rs2515797	G	Women	No	-0.199 (-0.271,-0.127)	6.0E-08	0.014 (-0.057,0.085)	6.9E-01	3.6E-05
ECE1.3611.70.4	rs116386809	G	Men	No	-0.056 (-0.262,0.150)	6.0E-01	-0.638 (-0.873,-0.404)	9.7E-08	2.6E-04
ECE1.8767.44.3	rs11047985	C	Men	No	-0.023 (-0.095,0.048)	5.2E-01	0.183 (0.117,0.249)	5.2E-08	3.2E-05

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
EFNA3.5759.10.3	rs58769948	G	Women	No	0.297 (0.188,0.405)	8.4E-08	-0.091 (-0.198,0.017)	9.9E-02	6.9E-07
EFNB3.2514.65.3	rs17065165	A	Men	No	-0.165 (-0.244,-0.085)	4.7E-05	0.209 (0.133,0.285)	7.1E-08	2.5E-11
EGFR.5328.33.37	rs10468843	G	Men	No	-0.006 (-0.091,0.078)	8.8E-01	-0.233 (-0.318,-0.148)	7.0E-08	2.0E-04
EIF3G.11454.87.3	rs6565994	G	Women	No	-0.216 (-0.294,-0.138)	5.6E-08	0.024 (-0.055,0.104)	5.5E-01	2.4E-05
EIF4E2.9745.20.3	rs6576171	A	Women	No	0.240 (0.152,0.328)	8.7E-08	-0.065 (-0.153,0.023)	1.5E-01	1.5E-06
ELK1.10006.25.3	rs142806824	A	Women	No	0.199 (0.129,0.268)	2.1E-08	0.002 (-0.067,0.071)	9.5E-01	8.7E-05
ELMO1.12764.3.3	rs828842	G	Women	No	0.191 (0.121,0.260)	8.4E-08	-0.008 (-0.078,0.061)	8.1E-01	7.6E-05
ENO2.10339.48.3	rs34428701	T	Women	No	-0.307 (-0.418,-0.196)	6.5E-08	0.020 (-0.088,0.128)	7.2E-01	3.6E-05
EPHA1.3431.54.2	rs78237190	G	Women	Yes	0.507 (0.323,0.691)	6.5E-08	-0.129 (-0.313,0.055)	1.7E-01	1.6E-06
EPHA2.4834.61.2	rs911148	C	Women	Yes	-0.203 (-0.277,-0.129)	7.1E-08	0.033 (-0.040,0.106)	3.8E-01	8.7E-06
ERAP1.4964.67.1	rs6859054	T	Men	No	-0.000 (-0.089,0.089)	1.0e+00	-0.255 (-0.343,-0.167)	1.2E-08	6.4E-05
ERCC4.9895.77.3	rs113627905	C	Women	No	0.231 (0.151,0.311)	1.6E-08	0.009 (-0.069,0.088)	8.1E-01	1.0E-04
ERP29.13728.19.3	rs10950846	A	Men	Yes	-0.115 (-0.252,0.022)	1.0E-01	0.425 (0.294,0.557)	2.4E-10	2.5E-08
ERP44.6064.4.3	rs6074977	C	Women	No	-0.542 (-0.737,-0.346)	5.5E-08	0.051 (-0.148,0.250)	6.1E-01	3.0E-05
ERP44.6064.4.3	rs1012334	A	Women	No	0.200 (0.131,0.269)	1.2E-08	-0.006 (-0.076,0.063)	8.6E-01	3.4E-05
ESAM.2981.9.3	rs117400722	G	Men	No	0.046 (-0.154,0.247)	6.5E-01	-0.511 (-0.696,-0.325)	6.6E-08	6.4E-05
ESD.4984.83.1	rs4646391	C	Men	No	0.001 (-0.069,0.070)	9.9E-01	-0.194 (-0.262,-0.126)	1.9E-08	8.1E-05
ETS2.12350.86.3	rs57192414	T	Women	Yes	0.248 (0.171,0.324)	2.8E-10	-0.038 (-0.114,0.038)	3.2E-01	2.2E-07
EVA1C.8877.22.3	rs1983765	T	Women	No	0.210 (0.137,0.282)	1.3E-08	-0.047 (-0.118,0.024)	1.9E-01	6.5E-07
FAIM3.6574.11.3	rs61940590	G	Men	No	-0.092 (-0.163,-0.022)	1.0E-02	0.195 (0.126,0.263)	2.5E-08	9.9E-09
FAM134B.8556.5.3	rs2983713	T	Men	No	0.061 (-0.032,0.155)	2.0E-01	-0.271 (-0.369,-0.172)	7.0E-08	1.7E-06
FAM150B.6284.7.3	rs17289966	T	Men	No	0.058 (-0.076,0.191)	4.0E-01	-0.367 (-0.500,-0.234)	6.4E-08	1.0E-05
FAM151A.7856.51.3	rs582835	T	Men	No	-0.036 (-0.110,0.038)	3.4E-01	0.212 (0.139,0.286)	1.3E-08	2.9E-06
FAM163B.10880.38.3	rs10942276	G	Men	No	-0.075 (-0.231,0.081)	3.4E-01	0.422 (0.271,0.574)	4.5E-08	7.1E-06
FAM171B.8061.102.3	rs9472777	G	Men	No	0.007 (-0.073,0.086)	8.7E-01	-0.214 (-0.292,-0.136)	7.1E-08	1.0E-04

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
FAM171B.8786.6.3	rs114439777	C	Men	No	-0.052 (-0.138,0.034)	2.3E-01	0.246 (0.162,0.331)	1.2E-08	1.2E-06
FAM171B.8851.42.3	rs79585054	C	Women	No	0.442 (0.282,0.602)	5.9E-08	0.006 (-0.155,0.167)	9.4E-01	1.7E-04
FAM171B.8851.42.3	rs117153691	C	Men	No	0.064 (-0.114,0.241)	4.8E-01	-0.495 (-0.672,-0.318)	4.0E-08	1.2E-05
FAM174A.6597.24.3	rs4485089	A	Men	No	0.029 (-0.052,0.110)	4.9E-01	-0.227 (-0.305,-0.148)	1.4E-08	9.0E-06
FAM174A.6597.24.3	rs2723264	T	Women	No	0.235 (0.149,0.322)	9.5E-08	-0.011 (-0.097,0.076)	8.1E-01	8.1E-05
FAM19A4.6511.17.3	rs12500806	C	Men	Yes	0.006 (-0.078,0.089)	9.0E-01	-0.255 (-0.337,-0.174)	8.5E-10	1.2E-05
FAM213A.13423.94.3	rs7862702	C	Women	No	0.207 (0.131,0.283)	1.0E-07	-0.015 (-0.091,0.061)	7.0E-01	5.3E-05
FAM24B.8775.61.3	rs4938846	C	Women	No	0.343 (0.218,0.467)	7.2E-08	0.007 (-0.119,0.132)	9.2E-01	2.0E-04
FAS.9459.7.3	rs263332	T	Men	No	0.001 (-0.083,0.084)	9.9E-01	-0.226 (-0.306,-0.147)	2.3E-08	1.1E-04
FBLN1.6470.19.3	rs6552526	C	Men	No	0.031 (-0.103,0.165)	6.5E-01	-0.360 (-0.485,-0.236)	1.5E-08	2.7E-05
FBXO3.14628.72.3	rs1394564	A	Men	No	0.036 (-0.038,0.109)	3.4E-01	-0.202 (-0.276,-0.129)	7.1E-08	7.3E-06
FCGR2A.3309.2.2	rs73193587	T	Men	Yes	-0.017 (-0.091,0.057)	6.5E-01	0.197 (0.125,0.268)	8.3E-08	5.3E-05
FCGR2A.3309.2.2	rs6900370	A	Men	No	0.011 (-0.102,0.124)	8.5E-01	-0.297 (-0.404,-0.191)	4.7E-08	1.0E-04
FCRL5.6103.70.3	rs115230181	A	Women	Yes	-0.690 (-0.921,-0.458)	5.1E-09	0.054 (-0.172,0.280)	6.4E-01	6.6E-06
FCRL6.6617.12.3	rs7913390	C	Women	No	-0.209 (-0.283,-0.134)	3.9E-08	0.081 (0.006,0.156)	3.4E-02	7.4E-08
FGF4.4123.60.2	rs1031267	T	Women	No	-0.303 (-0.414,-0.192)	8.8E-08	-0.005 (-0.117,0.108)	9.4E-01	2.2E-04
FGF7.4487.1.1	rs10146530	C	Men	No	0.046 (-0.034,0.127)	2.6E-01	-0.214 (-0.290,-0.137)	3.9E-08	4.3E-06
FGF7.4487.1.1	rs150021587	G	Women	No	-0.596 (-0.815,-0.378)	8.4E-08	0.008 (-0.183,0.199)	9.4E-01	4.4E-05
FGFR4.4988.49.2	rs59231956	C	Women	No	-0.276 (-0.376,-0.176)	6.4E-08	0.055 (-0.040,0.150)	2.5E-01	2.6E-06
FKBP7.9288.7.3	rs3092288	G	Women	No	0.212 (0.134,0.289)	9.3E-08	0.001 (-0.074,0.076)	9.8E-01	1.3E-04
FKBP7.9288.7.3	rs28670583	A	Men	No	0.039 (-0.033,0.111)	2.9E-01	-0.197 (-0.266,-0.128)	2.2E-08	3.5E-06
FLRT3.9128.34.3	rs34145303	A	Men	No	0.068 (-0.078,0.214)	3.6E-01	-0.390 (-0.532,-0.247)	8.9E-08	1.1E-05
FLRT3.9128.34.3	rs882275	C	Men	No	0.010 (-0.060,0.080)	7.8E-01	-0.183 (-0.250,-0.116)	7.9E-08	9.0E-05
FOXJ2.14204.55.3	rs138184725	G	Men	No	0.002 (-0.216,0.220)	9.8E-01	0.632 (0.400,0.863)	9.2E-08	1.1E-04
FOXL2.11375.49.3	rs1411381	G	Men	No	-0.075 (-0.151,0.002)	5.7E-02	0.205 (0.131,0.279)	6.7E-08	3.0E-07

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
FST.4132.27.2	rs56359796	G	Men	No	0.010 (-0.066,0.087)	7.9E-01	0.211 (0.137,0.284)	2.2E-08	2.2E-04
FTH1.FTL.3031.66.1	rs75632028	A	Men	No	-0.072 (-0.185,0.041)	2.1E-01	0.298 (0.189,0.407)	7.9E-08	3.9E-06
FUT3.4548.4.2	rs12456689	T	Women	Yes	0.247 (0.159,0.334)	3.3E-08	-0.059 (-0.147,0.030)	2.0E-01	1.5E-06
FUT9.6991.24.3	rs336962	A	Women	No	0.199 (0.126,0.271)	7.1E-08	0.011 (-0.058,0.080)	7.6E-01	2.3E-04
G6B.8659.68.3	rs903836	T	Men	No	-0.074 (-0.266,0.117)	4.5E-01	0.510 (0.330,0.690)	2.8E-08	1.3E-05
GBP6.7818.101.3	rs10077981	A	Men	No	0.001 (-0.068,0.070)	9.7E-01	0.183 (0.116,0.249)	7.0E-08	2.0E-04
GCK.12960.9.3	rs12708701	T	Men	Yes	-0.006 (-0.109,0.097)	9.1E-01	-0.314 (-0.422,-0.206)	1.1E-08	5.2E-05
GCK.12960.9.3	rs573520313	T	Women	No	0.554 (0.369,0.739)	4.6E-09	0.031 (-0.153,0.215)	7.4E-01	8.6E-05
GFRA2.2515.14.3	rs116025596	C	Women	Yes	-0.685 (-0.930,-0.440)	4.1E-08	0.035 (-0.200,0.269)	7.7E-01	3.1E-05
GFRA2.2515.14.3	rs112216801	C	Men	No	0.008 (-0.075,0.090)	8.6E-01	0.253 (0.169,0.337)	3.9E-09	4.6E-05
GLIPR1.9265.10.3	rs7806154	T	Men	No	-0.006 (-0.114,0.101)	9.1E-01	-0.302 (-0.402,-0.201)	3.8E-09	8.0E-05
GLTP.12513.8.3	rs2718778	G	Women	No	-0.271 (-0.369,-0.172)	8.1E-08	0.003 (-0.095,0.100)	9.6E-01	1.1E-04
GNPNAT1.13954.9.3	rs80334360	A	Men	No	-0.008 (-0.115,0.099)	8.8E-01	0.305 (0.193,0.416)	8.9E-08	7.3E-05
GNPTG.10666.7.3	rs10066049	T	Women	No	0.270 (0.180,0.359)	3.5E-09	-0.003 (-0.090,0.084)	9.5E-01	1.8E-05
GNS.8314.71.3	rs7717453	C	Men	No	0.008 (-0.073,0.089)	8.5E-01	-0.214 (-0.290,-0.138)	3.0E-08	8.5E-05
GPD1L.12420.10.3	rs10077981	A	Men	No	-0.000 (-0.069,0.069)	9.9E-01	0.193 (0.126,0.259)	1.3E-08	7.9E-05
GPI.4272.46.2	rs11088998	G	Women	Yes	0.233 (0.148,0.318)	8.5E-08	-0.019 (-0.103,0.066)	6.6E-01	4.0E-05
GPI.4272.46.2	rs77546608	A	Men	No	0.035 (-0.171,0.240)	7.4E-01	-0.511 (-0.694,-0.328)	4.7E-08	1.0E-04
GPNMB.5080.131.3	rs72702815	T	Women	No	0.464 (0.298,0.630)	4.4E-08	-0.123 (-0.283,0.036)	1.3E-01	5.6E-07
GRAMD1C.8842.16.3	rs61077924	C	Women	No	-0.524 (-0.594,-0.454)	1.1E-48	-0.711 (-0.774,-0.649)	1.2E-110	8.9E-05
GRAMD1C.8842.16.3	rs4682146	C	Men	No	-0.523 (-0.594,-0.453)	1.1E-47	-0.719 (-0.781,-0.656)	9.6E-112	5.2E-05
GRP.5897.58.3	rs4817013	T	Women	No	0.194 (0.123,0.266)	9.8E-08	0.006 (-0.063,0.075)	8.6E-01	2.1E-04
GRP.5897.58.3	rs4872232	C	Women	No	-0.192 (-0.260,-0.123)	4.7E-08	-0.005 (-0.075,0.065)	8.9E-01	1.9E-04
GRP.8400.74.3	rs72892585	C	Men	No	-0.014 (-0.223,0.194)	8.9E-01	0.712 (0.502,0.922)	3.1E-11	1.5E-06
GSTA4.14645.253.3	rs33999362	A	Women	No	-0.195 (-0.265,-0.125)	4.9E-08	0.072 (0.003,0.140)	4.1E-02	1.0E-07

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
GSTK1.13474.40.3	rs62176181	G	Men	No	-0.015 (-0.131,0.101)	8.0E-01	0.307 (0.197,0.418)	5.0E-08	8.3E-05
GSTK1.13474.40.3	rs112603185	A	Men	No	0.147 (-0.035,0.330)	1.1E-01	-0.456 (-0.623,-0.289)	8.8E-08	1.8E-06
GUCA2B.6223.5.3	rs10887641	A	Women	No	-0.191 (-0.260,-0.121)	7.3E-08	0.038 (-0.030,0.106)	2.7E-01	4.0E-06
GZMH.3373.5.2	rs730943	A	Men	No	-0.010 (-0.083,0.062)	7.8E-01	0.198 (0.128,0.269)	3.8E-08	5.6E-05
H2AFZ.4163.5.2	rs7448318	C	Men	No	0.037 (-0.040,0.114)	3.5E-01	-0.201 (-0.275,-0.128)	8.6E-08	1.2E-05
HAVCR1.9021.1.3	rs12490183	C	Men	Yes	0.058 (-0.093,0.209)	4.5E-01	-0.418 (-0.564,-0.272)	1.9E-08	8.7E-06
HBEGF.14094.29.3	rs12490098	A	Men	No	-0.196 (-0.358,-0.033)	1.8E-02	0.459 (0.293,0.625)	6.1E-08	3.3E-08
HDGFRP2.4553.65.3	rs56175853	A	Women	No	-0.270 (-0.366,-0.173)	4.7E-08	0.016 (-0.080,0.112)	7.5E-01	4.0E-05
HECW1.12669.30.3	rs17636165	G	Women	No	-0.251 (-0.340,-0.162)	3.4E-08	-0.008 (-0.092,0.076)	8.6E-01	1.0E-04
HEPACAM2.6568.18.3	rs73133989	C	Men	No	0.134 (-0.026,0.295)	1.0E-01	-0.445 (-0.609,-0.282)	9.1E-08	7.1E-07
HEPACAM2.9116.28.3	rs160283	G	Men	No	0.038 (-0.043,0.120)	3.5E-01	-0.227 (-0.307,-0.146)	3.1E-08	5.3E-06
HEPACAM2.9116.28.3	rs13156973	A	Women	No	-0.217 (-0.290,-0.145)	4.6E-09	-0.004 (-0.076,0.067)	9.0E-01	4.2E-05
HEPACAM2.9116.28.3	rs75311297	G	Women	No	0.358 (0.228,0.488)	6.9E-08	-0.032 (-0.165,0.101)	6.4E-01	4.0E-05
HEPACAM2.9116.28.3	rs77451450	T	Men	No	0.015 (-0.098,0.128)	7.9E-01	-0.287 (-0.392,-0.183)	7.5E-08	1.2E-04
HMGB3.12775.6.3	rs344956	C	Women	No	-0.203 (-0.273,-0.133)	1.4E-08	-0.004 (-0.073,0.064)	9.0E-01	7.0E-05
HNRNPDL.10852.114.3	rs11088999	C	Women	Yes	0.214 (0.135,0.293)	1.0E-07	-0.030 (-0.108,0.048)	4.5E-01	1.6E-05
HNRNPM.12783.29.3	rs56292770	G	Women	Yes	0.217 (0.138,0.296)	6.9E-08	-0.032 (-0.109,0.046)	4.2E-01	1.0E-05
HNRNPM.12783.29.3	rs3132950	C	Men	Yes	-0.022 (-0.094,0.051)	5.6E-01	0.194 (0.125,0.263)	3.9E-08	2.4E-05
HOGA1.10024.44.3	rs29579	C	Men	No	0.048 (-0.041,0.137)	2.9E-01	-0.238 (-0.323,-0.152)	5.2E-08	5.6E-06
HRK.8008.28.3	rs7277498	A	Men	No	-0.005 (-0.082,0.073)	9.0E-01	-0.222 (-0.297,-0.146)	8.3E-09	8.5E-05
HS3ST3B1.6986.17.3	rs56311947	G	Men	No	0.005 (-0.066,0.076)	9.0E-01	-0.192 (-0.262,-0.122)	6.9E-08	1.1E-04
HTRA2.3317.33.1	rs4306167	G	Men	No	-0.011 (-0.087,0.065)	7.8E-01	0.206 (0.133,0.279)	3.0E-08	5.3E-05
ICAM5.8245.27.3	rs78706926	T	Women	No	0.320 (0.209,0.431)	1.7E-08	-0.033 (-0.149,0.083)	5.7E-01	1.6E-05
ICOSLG.9303.9.3	rs115053672	C	Men	No	-0.062 (-0.265,0.141)	5.5E-01	0.480 (0.303,0.656)	9.8E-08	7.7E-05
ID1.9436.2.3	rs680119	T	Women	No	0.197 (0.126,0.267)	4.9E-08	-0.007 (-0.077,0.064)	8.5E-01	6.3E-05

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
IDE.3197.70.2	rs13211315	G	Men	No	-0.025 (-0.098,0.047)	4.9E-01	0.200 (0.128,0.272)	5.1E-08	1.6E-05
IDO1.9759.13.3	rs12310980	G	Men	No	0.002 (-0.071,0.075)	9.5E-01	0.199 (0.127,0.271)	6.4E-08	1.7E-04
IFNA14.7180.114.3	rs17006826	T	Women	No	-0.209 (-0.282,-0.136)	2.1E-08	-0.012 (-0.083,0.058)	7.4E-01	1.5E-04
IFNG.14147.50.3	rs9295360	T	Men	No	0.003 (-0.068,0.074)	9.3E-01	0.188 (0.119,0.256)	8.8E-08	2.5E-04
IFNL1.13734.22.3	rs10067159	C	Men	No	-0.017 (-0.086,0.052)	6.2E-01	0.199 (0.132,0.266)	7.0E-09	1.1E-05
IGDCC3.11952.1.3	rs9267845	A	Men	No	-0.005 (-0.078,0.067)	8.8E-01	0.192 (0.122,0.261)	5.6E-08	1.2E-04
IGFBP1.13741.36.3	rs2050087	C	Women	Yes	0.252 (0.167,0.338)	8.2E-09	-0.100 (-0.187,-0.014)	2.2E-02	1.3E-08
IGFBPL1.7815.49.3	rs2662723	G	Men	No	-0.016 (-0.087,0.055)	6.5E-01	0.188 (0.121,0.255)	4.3E-08	4.2E-05
IGHE.IGK.IGL.4135.84.2	rs752626	C	Women	Yes	-0.225 (-0.308,-0.143)	7.5E-08	0.098 (0.018,0.178)	1.6E-02	3.2E-08
IGJ.8362.102.3	rs72769720	C	Women	No	-0.452 (-0.601,-0.303)	2.8E-09	-0.010 (-0.158,0.138)	8.9E-01	3.7E-05
IL12A.IL12B.10367.26.3	rs3851630	T	Men	No	-0.077 (-0.174,0.019)	1.2E-01	0.245 (0.157,0.334)	5.2E-08	1.3E-06
IL12A.IL12B.10367.26.3	rs79791086	T	Women	No	-0.546 (-0.737,-0.354)	2.5E-08	-0.008 (-0.202,0.185)	9.3E-01	1.1E-04
IL15RA.3445.53.2	rs9599203	A	Women	No	0.368 (0.237,0.499)	3.5E-08	-0.042 (-0.174,0.089)	5.3E-01	1.4E-05
IL17C.9255.5.3	rs144631133	A	Men	No	0.017 (-0.156,0.190)	8.5E-01	0.527 (0.354,0.700)	2.4E-09	4.3E-05
IL17RB.5084.154.3	rs77301136	C	Men	No	0.007 (-0.180,0.194)	9.4E-01	0.550 (0.362,0.738)	9.5E-09	5.9E-05
IL17RB.5084.154.3	rs75029903	C	Men	No	-0.008 (-0.145,0.128)	9.1E-01	0.421 (0.282,0.561)	3.2E-09	1.6E-05
IL17RC.5468.67.3	rs2267266	C	Men	No	0.019 (-0.053,0.090)	6.1E-01	-0.201 (-0.271,-0.131)	2.2E-08	1.7E-05
IL17RD.3376.49.2	rs35706962	C	Women	No	0.428 (0.283,0.572)	6.5E-09	0.010 (-0.132,0.153)	8.9E-01	5.6E-05
IL18R1.14079.14.3	rs892350	T	Men	No	-0.009 (-0.091,0.072)	8.2E-01	0.230 (0.151,0.309)	1.2E-08	3.6E-05
IL1R2.5666.64.3	rs12676148	C	Men	No	0.130 (0.004,0.255)	4.3E-02	-0.344 (-0.461,-0.226)	8.9E-09	6.7E-08
IL22.2778.10.2	rs1575962	A	Men	No	0.015 (-0.081,0.112)	7.5E-01	-0.256 (-0.349,-0.164)	5.9E-08	7.0E-05
IL22RA1.9603.9.3	rs3128390	C	Men	No	-0.033 (-0.109,0.043)	4.0E-01	0.218 (0.143,0.294)	1.6E-08	4.5E-06
IL22RA2.5087.5.3	rs78682114	C	Men	No	-0.058 (-0.199,0.084)	4.3E-01	0.408 (0.259,0.557)	8.3E-08	9.2E-06
IL23R.5088.175.3	rs7449190	T	Women	No	0.284 (0.186,0.383)	1.7E-08	-0.017 (-0.115,0.081)	7.4E-01	2.2E-05
IL25.4137.57.2	rs11103845	G	Women	Yes	-0.403 (-0.551,-0.255)	9.6E-08	0.089 (-0.060,0.238)	2.4E-01	4.5E-06

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
IL34.4556.10.2	rs17418632	A	Men	No	-0.088 (-0.191,0.014)	9.0E-02	0.288 (0.186,0.389)	2.8E-08	3.1E-07
IL5.3741.4.3	rs10981641	C	Men	No	0.007 (-0.141,0.155)	9.3E-01	0.406 (0.261,0.551)	4.4E-08	1.7E-04
ILF3.12759.47.3	rs7828742	A	Men	No	0.005 (-0.067,0.077)	8.9E-01	-0.199 (-0.271,-0.128)	4.2E-08	8.0E-05
ILF3.12759.47.3	rs1898489	T	Men	No	0.011 (-0.130,0.151)	8.8E-01	-0.366 (-0.499,-0.233)	7.3E-08	1.4E-04
IMPAD1.9231.23.3	rs11032153	G	Men	No	-0.005 (-0.175,0.165)	9.5E-01	-0.458 (-0.625,-0.290)	8.3E-08	2.0E-04
INSL3.5723.4.3	rs59037764	G	Women	No	0.221 (0.141,0.300)	5.7E-08	-0.013 (-0.091,0.065)	7.4E-01	4.1E-05
IRF1.10351.51.3	rs73173432	G	Women	No	-0.304 (-0.416,-0.192)	9.6E-08	0.091 (-0.015,0.197)	9.1E-02	4.8E-07
ISG15.14151.4.3	rs80038180	C	Women	No	-0.676 (-0.891,-0.460)	8.5E-10	-0.079 (-0.285,0.126)	4.5E-01	8.7E-05
ITGAL.11617.1.3	rs3928116	T	Men	No	-0.049 (-0.131,0.033)	2.4E-01	0.223 (0.144,0.302)	2.7E-08	2.7E-06
ITIH1.7955.195.3	rs9266806	T	Men	No	-0.009 (-0.095,0.076)	8.3E-01	-0.232 (-0.315,-0.149)	4.0E-08	2.5E-04
ITM2A.7765.15.3	rs543076146	C	Women	No	0.269 (0.174,0.364)	2.6E-08	-0.064 (-0.156,0.029)	1.8E-01	8.3E-07
ITPRIPL1.9221.6.3	rs56292770	G	Women	Yes	0.225 (0.146,0.304)	2.4E-08	-0.036 (-0.114,0.042)	3.6E-01	3.9E-06
ITSN1.14070.56.3	rs62580764	G	Women	Yes	-0.358 (-0.480,-0.237)	7.5E-09	0.064 (-0.047,0.175)	2.6E-01	5.2E-07
JAK2.4998.50.2	rs113065532	C	Women	No	0.451 (0.288,0.614)	5.8E-08	-0.036 (-0.192,0.119)	6.5E-01	2.2E-05
JAM2.2997.8.1	rs10182134	T	Men	No	-0.030 (-0.100,0.039)	3.9E-01	0.193 (0.126,0.261)	2.0E-08	6.4E-06
KAL1.6603.18.3	rs28733769	G	Men	No	-0.016 (-0.116,0.084)	7.6E-01	0.272 (0.174,0.369)	4.6E-08	5.5E-05
KCNE3.8067.21.3	rs35652435	G	Women	Yes	0.228 (0.150,0.306)	1.2E-08	-0.044 (-0.120,0.031)	2.5E-01	9.6E-07
KCNIP1.13650.11.3	rs2337632	A	Women	No	-0.216 (-0.295,-0.137)	9.8E-08	0.038 (-0.042,0.119)	3.5E-01	1.0E-05
KIAA1161.8093.13.3	rs16823123	A	Men	No	-0.051 (-0.242,0.141)	6.0E-01	0.475 (0.301,0.650)	9.4E-08	6.8E-05
KIAA1324.10637.50.3	rs4424962	C	Women	No	-0.249 (-0.341,-0.158)	8.3E-08	0.104 (0.017,0.192)	1.9E-02	4.1E-08
KIAA2013.6538.90.3	rs79181095	G	Men	No	0.032 (-0.087,0.151)	6.0E-01	-0.312 (-0.423,-0.202)	3.3E-08	3.2E-05
KIF16B.11672.17.3	rs146761139	C	Women	No	0.409 (0.261,0.558)	6.1E-08	-0.064 (-0.207,0.079)	3.8E-01	6.5E-06
KIR2DS2.10428.1.3	rs28886328	T	Women	No	0.360 (0.228,0.492)	9.4E-08	-0.075 (-0.197,0.048)	2.3E-01	2.3E-06
KIR3DS1.5097.14.3	rs1607567	G	Women	No	0.192 (0.121,0.262)	9.1E-08	0.002 (-0.068,0.072)	9.5E-01	1.8E-04
KIRREL3.4557.61.2	rs77274954	A	Men	No	-0.030 (-0.191,0.131)	7.2E-01	-0.469 (-0.635,-0.303)	3.1E-08	2.0E-04

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
KITLG.9377.25.3	rs2165346	T	Men	Yes	-0.073 (-0.161,0.015)	1.1E-01	0.234 (0.150,0.319)	4.9E-08	8.1E-07
KLK14.3681.87.3	rs11943860	G	Men	No	0.004 (-0.075,0.082)	9.3E-01	0.213 (0.136,0.289)	5.3E-08	1.9E-04
KLK14.8620.56.3	rs11059313	T	Women	No	-0.194 (-0.265,-0.123)	7.7E-08	0.027 (-0.044,0.098)	4.6E-01	1.5E-05
KLK3.8468.19.3	rs10993994	T	Men	Yes	-0.058 (-0.130,0.014)	1.2E-01	0.192 (0.124,0.261)	4.0E-08	8.7E-07
KLK3.8468.19.3	rs151224625	C	Men	Yes	-0.160 (-0.365,0.046)	1.3E-01	0.546 (0.350,0.741)	4.5E-08	1.1E-06
KLRC1.5629.58.3	rs67373475	C	Women	No	-0.247 (-0.337,-0.157)	7.5E-08	-0.008 (-0.098,0.083)	8.7E-01	2.3E-04
KLRC1.5629.58.3	rs6947549	T	Women	No	-0.310 (-0.414,-0.207)	4.6E-09	-0.001 (-0.100,0.099)	9.9E-01	2.4E-05
KMT2C.11402.17.3	rs79892278	G	Women	No	0.354 (0.229,0.478)	2.4E-08	-0.121 (-0.247,0.005)	5.9E-02	1.4E-07
KMT2C.11402.17.3	rs117992328	C	Men	No	-0.067 (-0.307,0.173)	5.8E-01	0.609 (0.393,0.826)	3.5E-08	4.1E-05
KPNA2.2860.19.2	rs62149015	G	Men	Yes	0.166 (0.027,0.306)	2.0E-02	-0.384 (-0.524,-0.244)	7.5E-08	4.9E-08
KPNA4.12698.72.3	rs7777594	G	Men	No	-0.041 (-0.231,0.149)	6.7E-01	0.519 (0.335,0.703)	3.5E-08	3.4E-05
KREMEN1.6512.68.3	rs35221840	G	Women	No	0.423 (0.267,0.579)	1.1E-07	-0.031 (-0.194,0.132)	7.1E-01	7.9E-05
KRTAP2.4.14615.46.3	rs317198	G	Men	No	0.012 (-0.112,0.135)	8.5E-01	0.345 (0.219,0.472)	8.9E-08	2.2E-04
LAG3.5099.14.3	rs2431501	T	Women	No	-0.214 (-0.286,-0.142)	6.2E-09	-0.017 (-0.090,0.055)	6.4E-01	1.7E-04
LAG3.9950.229.3	rs28437271	C	Women	No	-0.505 (-0.686,-0.323)	5.0E-08	0.072 (-0.101,0.244)	4.2E-01	6.3E-06
LAIR1.11284.24.3	rs793715	T	Women	No	0.506 (0.331,0.681)	1.4E-08	-0.064 (-0.227,0.098)	4.4E-01	2.8E-06
LAMC2.9580.5.3	rs112783675	G	Women	No	-0.502 (-0.655,-0.348)	1.4E-10	-0.048 (-0.201,0.104)	5.3E-01	4.0E-05
LAT2.5613.75.3	rs2277329	T	Women	No	-0.230 (-0.314,-0.145)	9.3E-08	0.035 (-0.046,0.116)	4.0E-01	9.2E-06
LCN2.2836.68.2	rs6871004	G	Women	No	0.297 (0.189,0.406)	7.6E-08	-0.009 (-0.118,0.101)	8.8E-01	1.0E-04
LCT.9017.58.3	rs55638935	C	Men	No	-0.007 (-0.115,0.100)	9.0E-01	-0.301 (-0.406,-0.197)	1.6E-08	1.2E-04
LDLR.13129.40.3	rs58124292	G	Men	No	0.003 (-0.085,0.092)	9.4E-01	0.256 (0.168,0.343)	1.1E-08	7.2E-05
LDLR.13129.40.3	rs58096832	G	Women	No	0.274 (0.174,0.375)	8.4E-08	-0.020 (-0.119,0.080)	7.0E-01	4.5E-05
LECT1.9928.125.3	rs12620171	A	Men	Yes	0.045 (-0.033,0.124)	2.5E-01	-0.225 (-0.303,-0.148)	1.3E-08	1.4E-06
LEP.8484.24.3	rs17047302	G	Women	No	0.341 (0.222,0.459)	2.1E-08	0.025 (-0.094,0.145)	6.8E-01	2.5E-04
LGALS8.4909.68.1	rs9587769	G	Men	No	-0.000 (-0.088,0.087)	9.9E-01	0.246 (0.160,0.332)	2.3E-08	8.3E-05

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
LGALS8.4909.68.1	rs6103273	G	Men	No	0.018 (-0.055,0.090)	6.3E-01	-0.195 (-0.265,-0.124)	6.3E-08	3.9E-05
LGALS8.4909.68.1	rs77189927	T	Men	No	0.004 (-0.161,0.170)	9.6E-01	-0.464 (-0.634,-0.294)	9.2E-08	1.1E-04
LGI3.8003.57.3	rs10014926	T	Women	No	-0.430 (-0.576,-0.283)	9.5E-09	0.198 (0.050,0.346)	8.7E-03	3.5E-09
LHB.8376.25.4	rs79609266	G	Men	No	-0.097 (-0.245,0.051)	2.0E-01	0.429 (0.272,0.587)	9.8E-08	1.8E-06
LILRA4.8299.66.3	rs78529238	T	Men	No	-0.024 (-0.228,0.181)	8.2E-01	0.557 (0.357,0.757)	4.5E-08	6.9E-05
LILRA4.8299.66.3	rs73283849	C	Men	No	-0.121 (-0.253,0.011)	7.3E-02	0.375 (0.242,0.509)	3.8E-08	2.3E-07
LILRB2.5091.28.3	rs4469949	G	Women	No	0.198 (0.126,0.270)	5.6E-08	0.007 (-0.066,0.081)	8.4E-01	2.6E-04
LILRB4.6453.70.3	rs2834904	C	Men	No	0.007 (-0.070,0.084)	8.6E-01	0.221 (0.146,0.297)	1.0E-08	9.8E-05
LINGO1.6620.82.3	rs35360713	G	Men	No	-0.001 (-0.109,0.108)	9.9E-01	-0.289 (-0.395,-0.183)	8.6E-08	1.9E-04
LINGO3.10827.67.3	rs17739723	G	Men	No	-0.024 (-0.113,0.065)	6.0E-01	0.239 (0.153,0.325)	4.5E-08	3.0E-05
LONP1.6398.12.3	rs73226816	T	Women	Yes	0.281 (0.180,0.382)	4.9E-08	-0.030 (-0.131,0.071)	5.6E-01	1.9E-05
LPCAT2.11175.45.3	rs981138	C	Men	Yes	0.114 (0.034,0.193)	4.9E-03	-0.219 (-0.296,-0.142)	2.7E-08	3.8E-09
LRFN1.7910.41.3	rs139297576	A	Men	No	-0.069 (-0.316,0.178)	5.9E-01	0.602 (0.385,0.818)	5.0E-08	6.3E-05
LRFN2.7200.4.3	rs12890474	A	Women	No	0.248 (0.157,0.340)	1.0E-07	-0.014 (-0.099,0.070)	7.4E-01	3.6E-05
LRIT2.11716.28.3	rs6877907	T	Women	No	-0.516 (-0.696,-0.336)	1.9E-08	0.049 (-0.120,0.217)	5.7E-01	7.2E-06
LRP1.10699.52.3	rs4925433	T	Men	Yes	0.069 (-0.048,0.186)	2.5E-01	-0.323 (-0.433,-0.212)	1.1E-08	1.9E-06
LRP1B.11275.94.3	rs6675466	G	Men	No	0.001 (-0.154,0.156)	9.9E-01	-0.425 (-0.580,-0.270)	7.4E-08	1.4E-04
LRPAP1.3640.14.3	rs61880881	A	Men	No	0.142 (-0.013,0.296)	7.2E-02	-0.360 (-0.491,-0.229)	7.5E-08	1.2E-06
LRPAP1.3640.14.3	rs17438440	C	Women	No	0.296 (0.189,0.402)	5.9E-08	0.001 (-0.107,0.108)	9.9E-01	1.4E-04
LRPAP1.8829.4.3	rs73161818	C	Men	No	-0.020 (-0.107,0.066)	6.5E-01	0.226 (0.143,0.309)	8.5E-08	5.5E-05
LRRC3.6917.49.3	rs72764980	C	Women	No	-0.631 (-0.853,-0.408)	2.7E-08	0.036 (-0.175,0.248)	7.3E-01	2.0E-05
LRRC37A2.8897.3.3	rs4143740	T	Women	No	-0.202 (-0.274,-0.131)	2.7E-08	0.046 (-0.024,0.116)	1.9E-01	1.1E-06
LRRC4C.9369.174.3	rs114714046	A	Women	Yes	-0.628 (-0.859,-0.397)	9.9E-08	0.132 (-0.102,0.367)	2.7E-01	6.0E-06
LRRN1.11293.14.3	rs11996097	G	Women	No	-0.635 (-0.860,-0.409)	3.3E-08	0.051 (-0.175,0.278)	6.6E-01	2.6E-05
LRRTM2.6904.14.3	rs3810459	A	Women	No	0.191 (0.121,0.261)	1.0E-07	0.006 (-0.063,0.075)	8.7E-01	2.3E-04

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
LRRTM4.6572.10.3	rs45624432	C	Women	No	0.472 (0.304,0.640)	3.9E-08	0.006 (-0.170,0.182)	9.5E-01	1.7E-04
LRRTM4.8646.61.3	rs4078399	A	Women	No	0.290 (0.198,0.381)	4.6E-10	-0.009 (-0.099,0.080)	8.4E-01	4.5E-06
LRTM1.9368.64.3	rs35846682	C	Women	No	0.335 (0.215,0.455)	4.7E-08	-0.011 (-0.128,0.106)	8.6E-01	5.3E-05
LTA.4703.87.2	rs7841	C	Men	No	0.016 (-0.102,0.133)	7.9E-01	0.326 (0.216,0.436)	6.1E-09	1.5E-04
LY9.3324.51.1	rs55993474	A	Women	No	0.451 (0.286,0.615)	7.7E-08	-0.035 (-0.195,0.125)	6.7E-01	3.3E-05
LYSMD4.9106.87.3	rs6973789	G	Women	No	0.420 (0.270,0.570)	4.5E-08	0.031 (-0.114,0.175)	6.8E-01	2.6E-04
MAGEA10.13610.9.3	rs17244095	C	Men	No	0.029 (-0.066,0.123)	5.5E-01	-0.250 (-0.341,-0.160)	6.3E-08	3.0E-05
MAGEB10.11456.2.3	rs2365650	T	Men	Yes	0.104 (-0.013,0.220)	8.1E-02	-0.308 (-0.418,-0.199)	3.4E-08	4.4E-07
MAGEB10.11456.2.3	rs13096279	G	Men	Yes	-0.105 (-0.245,0.035)	1.4E-01	0.386 (0.246,0.526)	7.2E-08	1.2E-06
MAN1A2.9077.10.3	rs10407041	T	Women	No	0.203 (0.129,0.277)	7.5E-08	0.006 (-0.068,0.080)	8.7E-01	2.2E-04
MAP2K4.5242.37.3	rs6833765	C	Women	No	0.323 (0.207,0.439)	4.7E-08	-0.020 (-0.135,0.095)	7.4E-01	4.0E-05
MAP3K3.12990.39.3	rs56292770	G	Women	Yes	0.224 (0.145,0.303)	2.7E-08	-0.013 (-0.091,0.064)	7.3E-01	2.7E-05
MEGF10.11168.3.3	rs273701	C	Women	No	0.200 (0.131,0.270)	1.8E-08	-0.016 (-0.085,0.053)	6.5E-01	1.5E-05
MEPE.3209.69.1	rs6467467	G	Women	No	-0.249 (-0.339,-0.158)	7.2E-08	0.010 (-0.083,0.103)	8.3E-01	9.2E-05
MET.2837.3.2	rs4641389	C	Women	Yes	-0.190 (-0.259,-0.121)	7.3E-08	0.019 (-0.048,0.086)	5.9E-01	2.3E-05
METTL1.12514.16.3	rs35157786	A	Women	No	-0.207 (-0.283,-0.131)	1.0E-07	0.021 (-0.055,0.096)	5.9E-01	3.2E-05
METTL1.12514.16.3	rs34875393	G	Women	Yes	-0.247 (-0.337,-0.157)	8.4E-08	0.098 (0.001,0.195)	4.8E-02	3.5E-07
MFAP1.5606.24.3	rs4770279	A	Men	No	-0.004 (-0.081,0.073)	9.2E-01	0.229 (0.154,0.304)	2.1E-09	2.1E-05
MFAP1.5606.24.3	rs7627256	C	Men	No	0.056 (-0.016,0.128)	1.3E-01	-0.218 (-0.288,-0.149)	6.6E-10	7.1E-08
MFNG.5605.77.3	rs28538862	G	Women	No	0.228 (0.146,0.310)	4.9E-08	-0.021 (-0.105,0.062)	6.1E-01	2.8E-05
MGAT4C.7208.60.3	rs17224975	T	Men	No	-0.011 (-0.161,0.138)	8.8E-01	-0.395 (-0.538,-0.252)	6.0E-08	2.7E-04
MILR1.11173.29.3	rs35648408	G	Women	No	0.221 (0.141,0.301)	6.5E-08	-0.021 (-0.101,0.059)	6.0E-01	2.7E-05
MLL2.13623.4.3	rs62049594	C	Men	No	0.033 (-0.036,0.101)	3.5E-01	-0.185 (-0.252,-0.118)	5.7E-08	8.1E-06
MMP10.3743.1.2	rs13052751	C	Men	No	0.009 (-0.065,0.082)	8.1E-01	-0.203 (-0.273,-0.132)	1.6E-08	4.7E-05
MMP16.9719.145.3	rs146137212	A	Men	No	-0.043 (-0.166,0.081)	5.0E-01	0.318 (0.204,0.431)	4.0E-08	2.5E-05

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
MNX1.9576.58.3	rs1967318	G	Men	No	0.000 (-0.072,0.072)	1.0e+00	0.186 (0.118,0.255)	9.1E-08	2.2E-04
MPG.12438.127.3	rs13021736	C	Men	No	-0.015 (-0.086,0.057)	6.9E-01	-0.202 (-0.272,-0.131)	2.1E-08	2.7E-04
MPZ.10615.18.3	rs12692367	C	Men	No	0.046 (-0.033,0.125)	2.5E-01	-0.219 (-0.296,-0.143)	2.0E-08	2.4E-06
MRAP.7895.108.3	rs1541837	A	Women	No	-0.199 (-0.270,-0.128)	4.5E-08	0.004 (-0.067,0.075)	9.1E-01	7.9E-05
MRC1.2637.77.2	rs6578065	T	Women	Yes	0.344 (0.219,0.469)	7.0E-08	-0.079 (-0.206,0.048)	2.2E-01	3.2E-06
MRPL52.7123.25.3	rs2292477	C	Men	Yes	-0.112 (-0.239,0.015)	8.5E-02	0.361 (0.231,0.491)	5.6E-08	3.6E-07
MSLN.3893.64.1	rs115135029	T	Men	No	-0.026 (-0.174,0.122)	7.3E-01	0.416 (0.270,0.561)	2.1E-08	3.0E-05
MSMB.10620.21.3	rs117036042	C	Men	No	-0.085 (-0.248,0.079)	3.1E-01	0.466 (0.300,0.633)	3.9E-08	3.7E-06
MTMR1.11167.6.3	rs61906909	G	Women	No	0.232 (0.148,0.316)	6.3E-08	0.010 (-0.072,0.092)	8.1E-01	2.1E-04
MTMR1.11167.6.3	rs17022806	T	Men	No	-0.006 (-0.152,0.139)	9.3E-01	0.380 (0.244,0.516)	4.6E-08	1.5E-04
MYCT1.13541.1.3	rs10079672	G	Women	No	-0.261 (-0.355,-0.167)	5.6E-08	0.034 (-0.066,0.135)	5.0E-01	2.6E-05
MYRF.8843.34.3	rs10478218	A	Women	No	0.340 (0.215,0.465)	9.4E-08	-0.067 (-0.187,0.054)	2.8E-01	4.4E-06
MYZAP.9055.81.3	rs7777594	G	Men	No	-0.102 (-0.291,0.088)	2.9E-01	0.502 (0.317,0.686)	9.8E-08	7.8E-06
N6AMT1.11096.57.3	rs80334360	A	Men	No	-0.025 (-0.132,0.081)	6.4E-01	0.321 (0.209,0.432)	1.7E-08	1.1E-05
NAB2.13680.3.3	rs2741088	G	Men	No	-0.008 (-0.106,0.090)	8.8E-01	-0.299 (-0.393,-0.206)	3.1E-10	2.4E-05
NAPB.12655.30.3	rs146477950	T	Men	No	-0.105 (-0.292,0.082)	2.7E-01	0.534 (0.353,0.716)	8.0E-09	1.5E-06
NCAM2.6507.16.3	rs577380	T	Women	No	0.196 (0.124,0.268)	9.6E-08	-0.000 (-0.075,0.074)	9.9E-01	2.0E-04
NCF2.10047.12.3	rs143589027	C	Women	No	0.312 (0.202,0.423)	3.0E-08	-0.002 (-0.114,0.109)	9.7E-01	8.5E-05
NCOA2.14045.12.3	rs995811	C	Men	No	0.030 (-0.048,0.108)	4.5E-01	-0.206 (-0.281,-0.132)	6.3E-08	1.8E-05
NCR1.8360.169.3	rs4696933	C	Men	No	0.002 (-0.087,0.090)	9.7E-01	-0.244 (-0.328,-0.160)	1.4E-08	8.5E-05
NDRG4.7934.11.3	rs187500955	C	Women	Yes	-0.545 (-0.738,-0.352)	3.3E-08	0.081 (-0.117,0.278)	4.2E-01	9.0E-06
NDST1.6927.7.3	rs10870218	C	Men	No	-0.024 (-0.094,0.047)	5.1E-01	0.192 (0.122,0.263)	8.6E-08	2.2E-05
NDUFB4.10677.9.3	rs11791803	T	Men	No	-0.017 (-0.088,0.055)	6.5E-01	0.186 (0.117,0.254)	1.0E-07	6.2E-05
NEFH.9900.36.3	rs11256342	T	Men	No	-0.022 (-0.099,0.056)	5.8E-01	-0.232 (-0.310,-0.153)	7.0E-09	1.9E-04
NELL1.6544.33.3	rs2973926	A	Women	Yes	0.257 (0.166,0.349)	3.3E-08	-0.084 (-0.175,0.006)	6.9E-02	1.9E-07

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
NET1.14260.112.3	rs419301	C	Men	No	-0.044 (-0.119,0.031)	2.5E-01	0.198 (0.125,0.271)	1.0E-07	5.5E-06
NGF.5801.72.3	rs6686129	T	Women	No	-0.215 (-0.289,-0.141)	1.2E-08	0.009 (-0.065,0.083)	8.2E-01	2.7E-05
NHP2L1.13602.6.3	rs56292770	G	Women	Yes	0.236 (0.157,0.314)	4.6E-09	-0.029 (-0.107,0.049)	4.6E-01	2.7E-06
NMB.9321.400.3	rs2486962	G	Women	No	0.196 (0.127,0.264)	2.8E-08	-0.000 (-0.069,0.068)	1.0e+00	7.9E-05
NME1.5909.51.3	rs7306771	T	Women	Yes	0.200 (0.128,0.271)	4.2E-08	-0.024 (-0.095,0.048)	5.1E-01	1.4E-05
NOG.5846.24.3	rs9367080	A	Men	No	-0.070 (-0.147,0.007)	7.2E-02	0.214 (0.137,0.292)	5.2E-08	3.0E-07
NOTUM.8252.2.3	rs76989619	C	Women	No	0.637 (0.403,0.871)	9.5E-08	0.025 (-0.189,0.238)	8.2E-01	1.5E-04
NOVA1.12557.18.3	rs338466	A	Women	No	-0.188 (-0.257,-0.119)	1.0E-07	0.019 (-0.048,0.086)	5.7E-01	2.5E-05
NPPA.5443.62.2	rs6789159	A	Women	No	-0.215 (-0.294,-0.136)	9.3E-08	-0.004 (-0.083,0.074)	9.1E-01	2.1E-04
NPTX2.6521.35.3	rs1859729	T	Men	Yes	-0.036 (-0.109,0.036)	3.3E-01	0.197 (0.128,0.266)	2.6E-08	5.2E-06
NPW.9986.14.3	rs2730927	C	Women	No	0.473 (0.306,0.640)	2.9E-08	-0.012 (-0.158,0.134)	8.7E-01	1.8E-05
NRG1.9178.30.3	rs149839436	G	Men	No	-0.036 (-0.251,0.178)	7.4E-01	0.585 (0.384,0.786)	1.3E-08	3.5E-05
NRG2.8060.7.3	rs4930806	A	Women	No	0.345 (0.224,0.467)	2.5E-08	-0.040 (-0.156,0.076)	5.0E-01	6.7E-06
NRG4.14139.16.3	rs1930594	C	Men	No	-0.030 (-0.120,0.059)	5.0E-01	0.233 (0.148,0.318)	8.3E-08	2.8E-05
NRP1.3214.3.2	rs7068721	C	Men	No	-0.020 (-0.096,0.056)	6.1E-01	0.204 (0.129,0.278)	7.7E-08	3.8E-05
NUDT16L1.12497.29.3	rs2277413	A	Women	Yes	0.498 (0.426,0.570)	2.7E-42	0.113 (0.037,0.189)	3.5E-03	4.4E-13
NUDT16L1.12497.29.3	rs10843767	T	Women	Yes	-0.194 (-0.262,-0.126)	2.2E-08	-0.004 (-0.075,0.067)	9.2E-01	1.4E-04
OBP2B.5680.54.3	rs2127855	C	Women	No	-0.343 (-0.467,-0.219)	6.1E-08	0.112 (-0.024,0.248)	1.1E-01	1.3E-06
OBP2B.5680.54.3	rs12196876	G	Women	Yes	0.266 (0.170,0.363)	6.8E-08	-0.029 (-0.123,0.065)	5.4E-01	1.7E-05
OCIAD1.3635.76.4	rs74396411	G	Women	No	0.391 (0.255,0.526)	1.6E-08	-0.116 (-0.254,0.022)	9.9E-02	2.7E-07
OLFML3.8660.5.3	rs16906040	C	Men	No	-0.008 (-0.186,0.170)	9.3E-01	0.509 (0.333,0.685)	1.6E-08	5.2E-05
ORC6L.12389.4.3	rs17116373	A	Men	No	0.025 (-0.135,0.186)	7.6E-01	-0.431 (-0.582,-0.281)	2.1E-08	4.8E-05
ORC6L.12389.4.3	rs11697735	G	Men	No	0.029 (-0.051,0.109)	4.8E-01	-0.228 (-0.303,-0.152)	3.0E-09	4.5E-06
OSM.14063.17.3	rs9360628	G	Women	No	-0.226 (-0.301,-0.152)	2.7E-09	0.020 (-0.055,0.094)	6.0E-01	4.6E-06
P2RX6.7233.73.3	rs4872232	C	Women	No	-0.199 (-0.267,-0.130)	1.4E-08	0.002 (-0.068,0.072)	9.6E-01	6.0E-05

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
PAIP1.12430.78.3	rs10962917	G	Men	No	-0.036 (-0.129,0.056)	4.4E-01	0.257 (0.168,0.346)	1.6E-08	7.5E-06
PAK3.3387.1.2	rs11047985	C	Men	No	-0.020 (-0.091,0.052)	5.9E-01	0.180 (0.114,0.246)	9.0E-08	5.9E-05
PAPSS1.14007.22.3	rs62271062	A	Women	No	0.658 (0.431,0.886)	1.5E-08	-0.196 (-0.395,0.003)	5.3E-02	3.0E-08
PARP11.12882.7.3	rs3759481	G	Men	No	0.035 (-0.080,0.151)	5.5E-01	-0.315 (-0.430,-0.200)	8.2E-08	2.5E-05
PATE4.8065.245.3	rs149437310	A	Women	No	-0.407 (-0.557,-0.257)	1.0E-07	0.085 (-0.062,0.233)	2.6E-01	4.5E-06
PATE4.8065.245.3	rs875500	T	Men	Yes	-0.005 (-0.083,0.072)	8.9E-01	0.841 (0.777,0.905)	5.2E-148	1.1E-61
PCBD2.6899.37.3	rs74849342	G	Women	Yes	0.505 (0.331,0.680)	1.4E-08	-0.076 (-0.236,0.084)	3.5E-01	1.5E-06
PCDHB1.9941.70.3	rs4800378	T	Men	No	-0.027 (-0.104,0.049)	4.9E-01	0.222 (0.149,0.294)	2.2E-09	3.7E-06
PCDHGC5.7983.1.3	rs12206986	A	Men	No	0.005 (-0.173,0.183)	9.5E-01	-0.484 (-0.655,-0.314)	2.7E-08	1.0E-04
PCOLCE2.6081.52.3	rs543258852	C	Men	Yes	0.300 (0.076,0.525)	8.8E-03	-0.575 (-0.775,-0.374)	2.0E-08	1.2E-08
PCSK1.13388.57.3	rs9322474	C	Women	Yes	-0.202 (-0.271,-0.133)	1.0E-08	0.009 (-0.060,0.077)	8.1E-01	2.3E-05
PDE6D.13491.40.3	rs678793	A	Women	No	0.331 (0.217,0.445)	1.4E-08	-0.029 (-0.134,0.075)	5.8E-01	5.2E-06
PDE9A.5201.50.4	rs144606797	C	Women	Yes	-0.420 (-0.569,-0.272)	2.9E-08	0.063 (-0.080,0.206)	3.9E-01	4.2E-06
PDGFRB.3459.49.2	rs6442329	G	Men	Yes	0.053 (-0.055,0.162)	3.3E-01	-0.308 (-0.413,-0.202)	1.2E-08	2.9E-06
PDIA3.7832.181.3	rs10486426	A	Men	Yes	-0.120 (-0.256,0.016)	8.4E-02	0.390 (0.258,0.523)	7.9E-09	1.4E-07
PENK.9076.25.3	rs7650602	T	Women	No	-0.196 (-0.266,-0.126)	3.7E-08	0.009 (-0.060,0.078)	8.0E-01	4.2E-05
PES1.4267.81.3	rs17045246	G	Women	No	0.563 (0.367,0.759)	1.9E-08	-0.094 (-0.283,0.094)	3.3E-01	2.2E-06
PGM1.9173.21.3	rs8129637	T	Men	No	0.046 (-0.026,0.118)	2.1E-01	-0.200 (-0.271,-0.129)	3.9E-08	1.9E-06
PGRMC1.7863.50.3	rs6731777	A	Women	No	0.374 (0.244,0.505)	1.7E-08	-0.019 (-0.150,0.112)	7.8E-01	3.0E-05
PHF3.11544.39.3	rs17692625	A	Women	No	-0.335 (-0.451,-0.219)	1.7E-08	-0.035 (-0.147,0.077)	5.4E-01	2.7E-04
PIM1.5359.65.3	rs73121347	C	Women	No	0.383 (0.252,0.514)	1.0E-08	-0.011 (-0.149,0.126)	8.7E-01	4.6E-05
PIN4.12718.43.3	rs72829969	C	Women	No	0.594 (0.380,0.808)	5.2E-08	-0.031 (-0.227,0.165)	7.6E-01	2.4E-05
PIWIL1.12793.4.3	rs79537600	T	Women	No	0.606 (0.385,0.828)	8.2E-08	0.039 (-0.168,0.246)	7.1E-01	2.5E-04
PKLR.11203.97.3	rs7170123	T	Men	No	0.072 (-0.058,0.203)	2.8E-01	-0.361 (-0.493,-0.230)	7.0E-08	4.3E-06
PLA2G2E.13670.81.3	rs6944897	T	Women	No	-0.476 (-0.647,-0.305)	5.0E-08	0.001 (-0.182,0.183)	9.9E-01	1.9E-04

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
PLAU.4158.54.2	rs6590199	G	Women	No	0.270 (0.177,0.364)	1.4E-08	0.021 (-0.071,0.114)	6.5E-01	2.0E-04
PMEL.6472.40.3	rs541786053	G	Women	No	-0.530 (-0.718,-0.342)	3.4E-08	-0.019 (-0.216,0.178)	8.5E-01	2.4E-04
PMS2.11312.40.3	rs62319809	T	Men	No	-0.005 (-0.080,0.071)	9.0E-01	0.203 (0.129,0.278)	1.0E-07	1.2E-04
POGLUT1.6467.65.3	rs10911507	A	Women	No	-0.203 (-0.277,-0.129)	8.3E-08	0.038 (-0.034,0.111)	3.0E-01	5.4E-06
POMC.4890.10.1	rs10266911	T	Men	No	0.007 (-0.066,0.079)	8.5E-01	0.201 (0.130,0.273)	3.5E-08	1.8E-04
POR.2731.29.2	rs78721608	A	Women	No	-0.489 (-0.668,-0.310)	8.7E-08	0.096 (-0.083,0.276)	2.9E-01	6.0E-06
POSTN.6645.53.3	rs7586684	G	Women	No	-0.370 (-0.499,-0.240)	2.4E-08	0.072 (-0.060,0.204)	2.8E-01	2.9E-06
PPCDC.13996.16.3	rs13014084	A	Men	No	0.039 (-0.119,0.197)	6.3E-01	-0.403 (-0.543,-0.263)	1.7E-08	4.1E-05
PPP3R1.3657.74.3	rs2898791	C	Men	No	-0.017 (-0.088,0.054)	6.4E-01	0.199 (0.128,0.270)	3.8E-08	2.5E-05
PRDM1.14197.2.3	rs79916008	G	Men	No	-0.022 (-0.173,0.129)	7.8E-01	-0.483 (-0.654,-0.312)	3.2E-08	7.5E-05
PRIM1.13591.31.3	rs17733351	G	Men	No	-0.034 (-0.112,0.043)	3.9E-01	0.209 (0.134,0.285)	6.0E-08	1.1E-05
PRNP.6545.58.3	rs793715	T	Women	Yes	0.513 (0.339,0.688)	8.5E-09	0.022 (-0.141,0.184)	7.9E-01	5.4E-05
PROK1.2247.20.11	rs79097477	G	Men	No	-0.076 (-0.226,0.074)	3.2E-01	0.419 (0.273,0.566)	1.9E-08	3.6E-06
PRPSAP1.9478.69.3	rs112291783	C	Men	No	-0.097 (-0.259,0.066)	2.4E-01	0.432 (0.278,0.585)	3.6E-08	3.6E-06
PRSS1.3049.61.2	rs7758999	A	Women	Yes	0.190 (0.120,0.259)	9.2E-08	-0.057 (-0.127,0.013)	1.1E-01	9.7E-07
PSAPL1.8814.33.3	rs13223283	G	Men	Yes	-0.072 (-0.151,0.006)	7.3E-02	0.210 (0.135,0.285)	4.7E-08	3.8E-07
PSG6.6456.17.3	rs201911713	T	Women	No	0.587 (0.373,0.802)	8.0E-08	-0.064 (-0.283,0.155)	5.7E-01	3.1E-05
PSG8.6238.55.3	rs2319921	C	Women	No	-0.263 (-0.347,-0.178)	1.1E-09	-0.009 (-0.097,0.079)	8.4E-01	4.4E-05
PSG9.9335.28.3	rs370045443	T	Women	No	0.367 (0.232,0.503)	1.0E-07	0.012 (-0.120,0.145)	8.5E-01	2.4E-04
PTH1R.13470.43.3	rs17428027	A	Men	No	0.002 (-0.073,0.078)	9.5E-01	-0.197 (-0.269,-0.124)	1.0E-07	1.9E-04
PTH2.7257.18.3	rs11073564	C	Women	Yes	-0.208 (-0.284,-0.133)	5.5E-08	0.044 (-0.029,0.117)	2.4E-01	2.5E-06
PTH2.7257.18.3	rs71446071	T	Men	Yes	0.169 (-0.008,0.347)	6.2E-02	-0.487 (-0.666,-0.308)	9.6E-08	3.4E-07
PTK2B.8918.64.3	rs75212570	C	Men	No	0.005 (-0.106,0.117)	9.2E-01	-0.322 (-0.426,-0.218)	1.3E-09	2.6E-05
PTPRD.9296.15.3	rs11788900	A	Men	No	-0.017 (-0.203,0.169)	8.6E-01	0.495 (0.314,0.676)	8.3E-08	1.1E-04
PTS.12014.19.3	rs139331832	A	Men	Yes	0.198 (0.026,0.371)	2.4E-02	-0.500 (-0.681,-0.320)	5.4E-08	4.0E-08

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
PTS.12014.19.3	rs17606027	T	Women	No	-0.301 (-0.410,-0.193)	5.0E-08	0.031 (-0.078,0.140)	5.8E-01	2.3E-05
PVRL2.6245.4.3	rs10145554	A	Women	No	0.213 (0.135,0.291)	9.0E-08	0.010 (-0.067,0.086)	8.0E-01	2.7E-04
PVRL3.13557.3.3	rs560049307	A	Men	No	0.018 (-0.125,0.160)	8.1E-01	0.440 (0.295,0.585)	2.6E-09	4.6E-05
PYGL.11441.11.3	rs10909972	C	Women	Yes	-0.338 (-0.462,-0.213)	1.1E-07	0.130 (-0.002,0.262)	5.4E-02	4.6E-07
QPCTL.8866.53.3	rs7965184	G	Women	No	0.196 (0.125,0.266)	5.6E-08	-0.002 (-0.074,0.070)	9.6E-01	1.3E-04
QPCTL.8866.53.3	rs75931502	C	Men	No	0.017 (-0.053,0.087)	6.4E-01	-0.193 (-0.264,-0.123)	8.4E-08	3.4E-05
QRFP.6463.59.3	rs1879159	C	Men	No	0.077 (-0.004,0.158)	6.2E-02	-0.226 (-0.307,-0.145)	4.4E-08	2.2E-07
QRFP.6463.59.3	rs116116091	G	Men	No	0.001 (-0.090,0.091)	9.8E-01	0.261 (0.168,0.353)	3.2E-08	8.3E-05
RAB14.14283.12.3	rs6864389	T	Women	No	-0.219 (-0.299,-0.139)	6.7E-08	0.051 (-0.026,0.128)	1.9E-01	1.7E-06
RAB27B.13596.3.3	rs4147531	G	Men	No	-0.036 (-0.104,0.033)	3.1E-01	0.186 (0.120,0.253)	3.4E-08	5.0E-06
RAD51L3.12554.10.3	rs28820590	C	Men	Yes	-0.101 (-0.184,-0.019)	1.6E-02	-0.423 (-0.503,-0.343)	2.3E-25	4.1E-08
RAP2A.9885.41.3	rs17247004	A	Men	No	-0.006 (-0.092,0.080)	8.8E-01	0.228 (0.146,0.310)	5.4E-08	1.1E-04
RAP2A.9885.41.3	rs185340883	C	Women	No	0.589 (0.376,0.802)	5.9E-08	-0.028 (-0.233,0.177)	7.9E-01	4.3E-05
RBL1.12879.5.3	rs2413588	A	Women	No	0.204 (0.129,0.280)	1.0E-07	-0.004 (-0.077,0.069)	9.2E-01	1.0E-04
RBM19.11468.15.3	rs77962070	G	Men	No	0.043 (-0.130,0.215)	6.3E-01	-0.493 (-0.666,-0.320)	2.3E-08	1.7E-05
RBM24.11380.84.3	rs370225926	G	Women	No	0.192 (0.124,0.261)	4.3E-08	-0.023 (-0.092,0.045)	5.1E-01	1.4E-05
RBM3.12747.89.3	rs283454	A	Women	Yes	-0.200 (-0.270,-0.131)	1.5E-08	-0.010 (-0.079,0.059)	7.7E-01	1.4E-04
RBP7.14208.3.3	rs9267845	A	Men	Yes	-0.002 (-0.075,0.070)	9.5E-01	0.201 (0.132,0.270)	1.1E-08	6.4E-05
REG3A.9277.16.3	rs34165183	C	Women	No	-0.329 (-0.449,-0.209)	7.7E-08	0.067 (-0.045,0.179)	2.4E-01	2.2E-06
RELB.11464.9.3	rs1935667	G	Women	No	-0.246 (-0.336,-0.157)	7.2E-08	0.042 (-0.045,0.129)	3.4E-01	6.1E-06
RELL2.6600.70.3	rs367621886	T	Women	No	0.219 (0.139,0.300)	9.6E-08	-0.025 (-0.105,0.056)	5.5E-01	2.7E-05
RELT.14112.40.3	rs73482001	C	Men	No	-0.017 (-0.090,0.057)	6.6E-01	-0.226 (-0.297,-0.155)	5.3E-10	5.9E-05
REPIN1.13554.78.3	rs188974339	T	Men	No	-0.016 (-0.170,0.139)	8.4E-01	-0.449 (-0.605,-0.293)	1.8E-08	1.1E-04
RET.3220.40.2	rs6853943	A	Men	No	-0.021 (-0.116,0.073)	6.6E-01	0.257 (0.162,0.352)	1.0E-07	4.4E-05
RETN.3046.31.1	rs1357302	T	Women	No	-0.200 (-0.273,-0.127)	8.5E-08	-0.003 (-0.075,0.070)	9.5E-01	1.7E-04

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
RFESD.13603.7.3	rs144265151	T	Women	No	-0.311 (-0.423,-0.200)	4.6E-08	0.059 (-0.051,0.169)	2.9E-01	3.7E-06
RFXAP.12726.3.3	rs111769779	G	Women	No	0.423 (0.276,0.569)	1.5E-08	-0.046 (-0.187,0.095)	5.3E-01	6.2E-06
RGS10.11634.32.3	rs112603185	A	Men	Yes	0.104 (-0.079,0.286)	2.7E-01	-0.455 (-0.622,-0.287)	9.9E-08	9.9E-06
RHOD.12442.4.3	rs117842273	C	Women	No	-0.499 (-0.676,-0.323)	2.9E-08	0.002 (-0.156,0.160)	9.8E-01	3.4E-05
RHOG.12540.25.3	rs6948789	C	Men	No	-0.026 (-0.110,0.058)	5.5E-01	0.230 (0.148,0.312)	4.4E-08	2.1E-05
RIPK2.8993.151.3	rs12668631	T	Women	No	-0.232 (-0.314,-0.149)	3.9E-08	0.060 (-0.023,0.143)	1.6E-01	1.1E-06
RNASE2.8394.56.3	rs4796505	C	Women	No	0.332 (0.213,0.451)	5.0E-08	0.027 (-0.084,0.138)	6.4E-01	2.5E-04
RNF148.7742.11.3	rs200866866	A	Men	No	-0.033 (-0.130,0.064)	5.1E-01	0.259 (0.165,0.352)	5.4E-08	2.2E-05
RNF43.14120.2.3	rs201435186	C	Women	No	-0.308 (-0.417,-0.198)	3.7E-08	-0.009 (-0.120,0.103)	8.8E-01	1.8E-04
RNGTT.12847.27.3	rs74717621	G	Women	No	-0.557 (-0.749,-0.366)	1.2E-08	-0.001 (-0.188,0.187)	1.0e+00	4.8E-05
RNPC3.12814.17.3	rs117174884	C	Women	Yes	-0.514 (-0.694,-0.334)	2.1E-08	0.170 (0.017,0.322)	2.9E-02	1.3E-08
RNPC3.12814.17.3	rs78487989	G	Men	Yes	0.093 (-0.025,0.211)	1.2E-01	-0.312 (-0.425,-0.198)	7.1E-08	1.2E-06
ROBO3.5117.14.3	rs11587874	T	Men	Yes	0.071 (0.001,0.140)	4.8E-02	-0.192 (-0.261,-0.122)	6.9E-08	1.8E-07
ROBO3.5117.14.3	rs316023	T	Men	No	0.005 (-0.071,0.081)	9.0E-01	0.202 (0.132,0.272)	1.7E-08	1.9E-04
ROCK2.13654.1.3	rs76279051	T	Men	No	0.006 (-0.102,0.115)	9.1E-01	0.292 (0.186,0.399)	7.8E-08	2.3E-04
ROR1.2590.69.4	rs6070721	T	Women	No	0.305 (0.194,0.416)	7.2E-08	0.015 (-0.091,0.122)	7.8E-01	2.3E-04
RPE.12646.2.3	rs11640887	C	Women	No	0.242 (0.165,0.320)	9.1E-10	-0.065 (-0.141,0.012)	9.9E-02	3.6E-08
RPL30.12478.15.3	rs56292770	G	Women	Yes	0.216 (0.137,0.295)	8.9E-08	-0.044 (-0.121,0.034)	2.7E-01	4.4E-06
RPS7.3864.5.2	rs56292770	G	Women	Yes	0.219 (0.140,0.298)	5.7E-08	-0.039 (-0.117,0.038)	3.2E-01	4.9E-06
RSPO3.13094.75.3	rs2054908	T	Women	Yes	0.231 (0.150,0.312)	2.2E-08	-0.053 (-0.133,0.028)	2.0E-01	1.1E-06
RSPO4.8464.31.3	rs11014312	A	Men	No	-0.077 (-0.188,0.035)	1.8E-01	0.290 (0.183,0.397)	1.0E-07	3.3E-06
RSPO4.8464.31.3	rs13325300	T	Women	Yes	-0.373 (-0.506,-0.241)	3.0E-08	0.078 (-0.056,0.212)	2.6E-01	2.6E-06
RXFP1.14135.3.3	rs4907192	T	Women	Yes	-0.436 (-0.591,-0.282)	2.9E-08	0.033 (-0.120,0.187)	6.7E-01	2.3E-05
S100A13.7223.60.3	rs77919600	A	Men	Yes	-0.146 (-0.351,0.058)	1.6E-01	0.559 (0.359,0.759)	4.3E-08	1.4E-06
S100A9.5339.49.3	rs3132950	C	Men	No	-0.022 (-0.095,0.050)	5.4E-01	0.208 (0.139,0.277)	3.4E-09	6.1E-06

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
SAMHD1.11303.7.3	rs13064030	A	Men	Yes	0.081 (-0.094,0.257)	3.6E-01	-0.487 (-0.667,-0.308)	1.0E-07	9.1E-06
SAP18.12830.4.3	rs2110690	A	Men	No	0.007 (-0.061,0.075)	8.4E-01	-0.197 (-0.264,-0.130)	7.2E-09	2.8E-05
SAP18.12830.4.3	rs2585132	G	Men	No	0.017 (-0.076,0.111)	7.1E-01	-0.237 (-0.323,-0.152)	5.9E-08	8.0E-05
SARS2.12348.46.3	rs62266384	C	Women	No	0.215 (0.143,0.288)	5.7E-09	-0.070 (-0.141,0.001)	5.3E-02	3.6E-08
SAYSD1.6594.64.3	rs79286296	C	Women	No	0.518 (0.332,0.704)	5.1E-08	0.005 (-0.177,0.187)	9.6E-01	1.1E-04
SCAMP5.13509.5.3	rs6467550	G	Women	No	0.251 (0.162,0.341)	3.5E-08	-0.024 (-0.113,0.065)	5.9E-01	1.9E-05
SCARB2.5100.53.3	rs7578883	A	Women	Yes	-0.260 (-0.339,-0.180)	1.5E-10	0.061 (-0.016,0.138)	1.2E-01	1.4E-08
SCARF2.5130.67.3	rs200884603	A	Women	No	0.391 (0.248,0.533)	7.6E-08	-0.039 (-0.180,0.101)	5.8E-01	2.5E-05
SCARF2.5130.67.3	rs6074977	C	Women	No	-0.546 (-0.742,-0.351)	4.3E-08	0.065 (-0.134,0.264)	5.2E-01	1.7E-05
SCARF2.9925.56.3	rs2660665	G	Men	No	0.045 (-0.026,0.115)	2.1E-01	-0.191 (-0.260,-0.122)	5.2E-08	2.7E-06
SCG3.7957.2.3	rs7654912	T	Women	No	-0.261 (-0.354,-0.169)	3.0E-08	0.003 (-0.090,0.097)	9.4E-01	7.7E-05
SCIN.12684.5.3	rs6472278	A	Women	No	-0.352 (-0.475,-0.229)	1.9E-08	0.051 (-0.077,0.178)	4.4E-01	8.2E-06
SCO1.7853.19.3	rs6679253	A	Men	Yes	0.094 (-0.003,0.191)	5.8E-02	-0.262 (-0.358,-0.166)	8.2E-08	3.1E-07
SEC13.14689.3.3	rs72794885	A	Men	Yes	-0.111 (-0.273,0.050)	1.8E-01	0.454 (0.289,0.620)	7.7E-08	1.6E-06
SEC61G.8872.1.3	rs1532330	T	Women	No	0.210 (0.134,0.287)	6.7E-08	-0.007 (-0.083,0.069)	8.6E-01	7.7E-05
SELP.4154.57.2	rs77628528	G	Men	No	-0.016 (-0.188,0.155)	8.5E-01	-0.465 (-0.623,-0.307)	8.1E-09	1.6E-04
SEMA3G.5628.21.3	rs7201599	C	Men	Yes	0.033 (-0.039,0.106)	3.7E-01	-0.204 (-0.277,-0.132)	2.8E-08	5.1E-06
SEMA4D.5737.61.3	rs4636031	A	Men	No	0.012 (-0.078,0.103)	7.9E-01	0.264 (0.173,0.355)	1.4E-08	1.2E-04
SERF1A.7842.52.3	rs10135970	A	Men	Yes	-0.252 (-0.428,-0.076)	5.0E-03	0.504 (0.327,0.680)	2.1E-08	2.8E-09
SERF1A.7842.52.3	rs4842885	A	Women	No	-0.418 (-0.571,-0.265)	8.4E-08	0.040 (-0.111,0.190)	6.0E-01	2.8E-05
SERPINA6.4785.30.3	rs9574691	C	Women	No	0.192 (0.122,0.262)	8.1E-08	-0.018 (-0.087,0.051)	6.1E-01	2.9E-05
SERPINA6.4785.30.3	rs74396411	G	Women	No	0.382 (0.246,0.518)	3.3E-08	0.003 (-0.135,0.141)	9.6E-01	1.2E-04
SERPINF2.3024.18.2	rs62119636	G	Women	No	-0.275 (-0.369,-0.181)	8.8E-09	-0.030 (-0.122,0.062)	5.2E-01	2.6E-04
SERPING1.4479.14.2	rs844138	G	Women	No	0.193 (0.124,0.261)	3.2E-08	-0.026 (-0.095,0.042)	4.5E-01	8.6E-06
SFRP1.3221.54.1	rs72711689	G	Women	Yes	-0.293 (-0.401,-0.186)	8.7E-08	0.031 (-0.070,0.133)	5.5E-01	1.7E-05

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
SFRP1.3221.54.1	rs12692019	A	Women	No	-0.248 (-0.335,-0.161)	2.1E-08	-0.003 (-0.086,0.081)	9.5E-01	6.3E-05
SFRP2.7751.121.3	rs3933269	T	Women	No	-0.410 (-0.552,-0.268)	1.5E-08	0.015 (-0.121,0.150)	8.3E-01	2.2E-05
SH3BP2.7769.29.3	rs11088998	G	Women	No	0.245 (0.160,0.330)	1.6E-08	0.009 (-0.075,0.094)	8.3E-01	1.2E-04
SIRPB1.6247.9.3	rs80329159	T	Women	No	0.532 (0.349,0.716)	1.3E-08	-0.025 (-0.192,0.142)	7.7E-01	1.1E-05
SIRT5.12461.8.3	rs4238111	C	Men	No	-0.030 (-0.112,0.051)	4.7E-01	0.214 (0.135,0.293)	1.0E-07	2.4E-05
SIT1.11194.6.3	rs7787886	G	Women	No	0.205 (0.130,0.280)	9.4E-08	-0.031 (-0.107,0.044)	4.2E-01	1.4E-05
SLAMF7.5487.7.3	rs6956539	T	Women	No	0.545 (0.354,0.737)	2.5E-08	-0.060 (-0.228,0.109)	4.9E-01	3.5E-06
SLC14A1.13430.50.3	rs3779034	C	Women	No	0.203 (0.133,0.274)	1.5E-08	0.004 (-0.066,0.075)	9.0E-01	9.4E-05
SLC35G2.13501.10.3	rs73052804	C	Men	Yes	-0.003 (-0.137,0.130)	9.6E-01	-0.358 (-0.487,-0.229)	5.7E-08	1.8E-04
SLC5A8.13691.10.3	rs28667269	A	Women	No	-0.200 (-0.271,-0.129)	3.1E-08	-0.001 (-0.072,0.071)	9.9E-01	9.7E-05
SLITRK4.7139.14.3	rs75854564	G	Women	No	-0.194 (-0.264,-0.125)	5.0E-08	0.004 (-0.066,0.074)	9.1E-01	8.0E-05
SLITRK4.7139.14.3	rs73937193	A	Women	No	-0.275 (-0.373,-0.177)	3.8E-08	-0.019 (-0.112,0.074)	6.9E-01	2.0E-04
SMC3.14324.52.3	rs2023222	T	Men	Yes	-0.007 (-0.079,0.066)	8.6E-01	0.193 (0.124,0.262)	4.5E-08	9.4E-05
SMCO2.7925.18.3	rs11210256	G	Men	No	-0.004 (-0.074,0.066)	9.0E-01	-0.202 (-0.270,-0.134)	5.6E-09	7.2E-05
SMIM9.8888.33.3	rs2515031	G	Men	No	0.007 (-0.064,0.078)	8.5E-01	-0.198 (-0.267,-0.128)	2.4E-08	5.5E-05
SNAI2.10014.31.3	rs10795292	T	Men	No	-0.004 (-0.114,0.105)	9.4E-01	-0.315 (-0.429,-0.200)	7.5E-08	1.2E-04
SNAI2.10014.31.3	rs75787785	C	Men	No	-0.009 (-0.217,0.199)	9.3E-01	0.519 (0.340,0.698)	1.4E-08	1.7E-04
SNAI2.10014.31.3	rs3920702	G	Women	No	0.196 (0.125,0.266)	5.2E-08	-0.104 (-0.173,-0.036)	2.9E-03	2.3E-09
SNCA.8458.16.3	rs111330842	C	Men	No	-0.090 (-0.204,0.025)	1.2E-01	0.294 (0.189,0.398)	3.6E-08	1.2E-06
SNRPA.12678.66.3	rs11088998	G	Women	Yes	0.234 (0.149,0.319)	7.3E-08	-0.035 (-0.120,0.049)	4.1E-01	1.1E-05
SOCS7.11657.86.3	rs5766023	C	Women	No	-0.193 (-0.263,-0.123)	7.4E-08	0.001 (-0.068,0.069)	9.8E-01	1.1E-04
SORBS3.12976.49.3	rs11038831	C	Women	Yes	-0.224 (-0.306,-0.143)	7.4E-08	0.060 (-0.018,0.139)	1.3E-01	8.6E-07
SPACA3.8076.4.3	rs137904288	G	Men	No	-0.052 (-0.235,0.131)	5.8E-01	0.463 (0.304,0.622)	1.2E-08	3.1E-05
SPAG11B.5762.35.3	rs1201976	A	Women	No	0.195 (0.125,0.265)	5.1E-08	-0.032 (-0.101,0.036)	3.6E-01	5.6E-06
SPARC.3043.49.2	rs9651726	T	Men	No	-0.010 (-0.131,0.112)	8.8E-01	0.330 (0.213,0.448)	3.7E-08	8.1E-05

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
SPARC.3043.49.2	rs116928246	G	Men	No	0.052 (-0.018,0.123)	1.5E-01	-0.210 (-0.279,-0.140)	3.9E-09	2.3E-07
SPARCL1.13707.27.3	rs852109	C	Men	No	0.006 (-0.100,0.112)	9.1E-01	-0.311 (-0.417,-0.204)	1.0E-08	3.6E-05
SPARCL1.4467.49.2	rs198805	G	Men	No	0.015 (-0.082,0.112)	7.6E-01	-0.274 (-0.369,-0.179)	1.5E-08	2.9E-05
SPINT1.2828.82.2	rs188628536	A	Men	Yes	-0.008 (-0.221,0.205)	9.4E-01	0.591 (0.378,0.805)	5.6E-08	9.7E-05
SPINT3.7926.13.3	rs6032259	A	Men	Yes	-0.161 (-0.230,-0.093)	3.9E-06	-0.823 (-0.878,-0.767)	4.0E-186	5.1E-49
SPOCK3.9906.21.3	rs1874278	A	Men	No	0.016 (-0.077,0.109)	7.4E-01	-0.238 (-0.325,-0.151)	8.3E-08	9.7E-05
SPOP.13727.44.3	rs67274280	G	Women	No	0.329 (0.218,0.440)	6.8E-09	0.015 (-0.100,0.130)	8.0E-01	1.2E-04
SPR.9287.6.3	rs3819091	C	Women	No	0.571 (0.371,0.770)	2.0E-08	0.025 (-0.183,0.232)	8.1E-01	2.0E-04
SPSB1.13942.140.3	rs10981641	C	Men	No	-0.036 (-0.184,0.112)	6.3E-01	0.399 (0.253,0.544)	7.7E-08	4.1E-05
SPTLC1.7886.26.3	rs62336320	G	Men	No	-0.086 (-0.165,-0.008)	3.1E-02	0.207 (0.131,0.283)	1.0E-07	1.4E-07
SSRP1.5032.64.1	rs6965978	A	Men	No	-0.030 (-0.108,0.049)	4.6E-01	0.222 (0.145,0.299)	1.5E-08	6.9E-06
ST6GALNAC3.10626.116.3	rs117220250	C	Women	No	-0.539 (-0.735,-0.343)	7.1E-08	0.076 (-0.099,0.252)	4.0E-01	4.6E-06
STAB1.14599.18.3	rs144793007	G	Women	No	0.454 (0.292,0.615)	3.7E-08	-0.082 (-0.256,0.092)	3.6E-01	9.9E-06
STAU1.12471.47.3	rs113462565	C	Women	No	0.361 (0.236,0.486)	1.5E-08	0.022 (-0.104,0.149)	7.3E-01	1.9E-04
STC2.6231.46.3	rs55970149	C	Men	No	-0.059 (-0.132,0.014)	1.2E-01	0.200 (0.127,0.273)	6.8E-08	8.8E-07
STC2.6231.46.3	rs10272049	C	Men	No	0.028 (-0.199,0.254)	8.1E-01	-0.657 (-0.894,-0.419)	6.0E-08	4.4E-05
STK17B.5249.31.3	rs6951678	C	Men	No	0.002 (-0.066,0.070)	9.4E-01	0.187 (0.119,0.254)	5.7E-08	1.6E-04
STMN2.10900.272.3	rs9267845	A	Men	No	-0.003 (-0.075,0.069)	9.4E-01	0.188 (0.119,0.257)	1.0E-07	1.9E-04
STX6.10945.11.3	rs12730100	A	Men	No	0.017 (-0.085,0.119)	7.5E-01	-0.300 (-0.400,-0.200)	4.2E-09	1.4E-05
SULT2A1.9829.91.3	rs10472000	A	Women	No	0.301 (0.196,0.406)	1.7E-08	0.003 (-0.109,0.114)	9.6E-01	1.3E-04
SUMF1.6941.11.3	rs115365541	C	Women	Yes	-0.583 (-0.798,-0.368)	1.0E-07	0.229 (0.026,0.431)	2.7E-02	7.1E-08
SUN5.11260.47.3	rs77806631	G	Women	No	-0.527 (-0.720,-0.333)	9.3E-08	0.122 (-0.079,0.323)	2.3E-01	5.0E-06
SVEP1.11109.56.3	rs75366433	G	Women	No	0.357 (0.226,0.489)	1.0E-07	-0.060 (-0.193,0.074)	3.8E-01	1.3E-05
SVEP1.11178.21.3	rs34159990	G	Men	No	0.016 (-0.079,0.111)	7.4E-01	-0.245 (-0.335,-0.155)	9.4E-08	9.6E-05
SVEP1.11178.21.3	rs75274627	G	Men	No	-0.010 (-0.162,0.142)	9.0E-01	-0.410 (-0.557,-0.263)	4.4E-08	2.0E-04

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
SYT11.7089.42.3	rs1012334	A	Women	No	0.194 (0.125,0.262)	3.9E-08	-0.018 (-0.087,0.052)	6.2E-01	2.4E-05
SYT11.7089.42.3	rs62336320	G	Men	No	-0.046 (-0.125,0.032)	2.5E-01	0.211 (0.135,0.288)	5.5E-08	3.8E-06
SYT5.9099.19.3	rs959065	A	Men	No	-0.014 (-0.087,0.058)	7.0E-01	0.190 (0.121,0.259)	7.7E-08	6.5E-05
TACSTD2.14034.22.3	rs77189383	G	Women	No	-0.458 (-0.620,-0.295)	3.4E-08	-0.039 (-0.191,0.113)	6.2E-01	2.3E-04
TACSTD2.14034.22.3	rs17651293	G	Women	No	0.271 (0.182,0.360)	2.1E-09	0.021 (-0.065,0.108)	6.3E-01	7.7E-05
TACSTD2.2619.72.2	rs288613	A	Men	No	0.001 (-0.074,0.076)	9.7E-01	-0.198 (-0.269,-0.126)	5.8E-08	1.7E-04
TAPBP.12378.71.3	rs56386479	C	Women	No	-0.454 (-0.605,-0.303)	3.8E-09	0.021 (-0.120,0.161)	7.7E-01	6.5E-06
TAX1BP1.12721.4.3	rs6886791	G	Women	No	0.323 (0.210,0.436)	2.1E-08	-0.097 (-0.209,0.015)	8.9E-02	2.2E-07
TAX1BP3.12498.12.3	rs115530557	T	Women	No	-0.577 (-0.786,-0.369)	5.4E-08	-0.014 (-0.231,0.203)	9.0E-01	2.4E-04
TEK.3773.15.4	rs113995269	C	Women	Yes	-0.367 (-0.500,-0.233)	7.2E-08	0.065 (-0.068,0.197)	3.4E-01	7.2E-06
TFF3.4721.54.2	rs9260306	C	Men	No	-0.043 (-0.191,0.104)	5.7E-01	-0.552 (-0.702,-0.403)	3.7E-13	2.0E-06
TFF3.4721.54.2	rs2524277	G	Men	No	-0.282 (-0.423,-0.140)	9.8E-05	-0.874 (-1.009,-0.738)	1.3E-36	3.2E-09
TFF3.8323.163.3	rs9260306	C	Men	No	-0.052 (-0.200,0.096)	4.9E-01	-0.446 (-0.595,-0.296)	5.7E-09	2.5E-04
TFF3.8323.163.3	rs2524277	G	Men	No	-0.196 (-0.338,-0.054)	6.7E-03	-0.754 (-0.891,-0.616)	5.0E-27	3.2E-08
TFRC.6895.1.3	rs564548360	C	Women	No	0.511 (0.333,0.689)	1.9E-08	-0.007 (-0.169,0.155)	9.3E-01	2.5E-05
TFRC.6895.1.3	rs17662532	T	Women	No	0.554 (0.374,0.735)	1.7E-09	0.034 (-0.154,0.223)	7.2E-01	9.4E-05
THPO.5947.90.3	rs75751545	A	Women	No	0.286 (0.183,0.389)	5.2E-08	-0.018 (-0.116,0.080)	7.2E-01	2.9E-05
THSD7A.11174.8.3	rs2359238	G	Women	No	0.195 (0.124,0.266)	6.9E-08	-0.047 (-0.115,0.021)	1.8E-01	1.4E-06
TINAGL1.11192.168.3	rs12445561	C	Women	No	-0.246 (-0.337,-0.156)	9.7E-08	-0.003 (-0.093,0.087)	9.4E-01	1.9E-04
TJP1.12001.7.3	rs6597545	G	Men	No	0.006 (-0.069,0.082)	8.7E-01	0.203 (0.130,0.276)	4.5E-08	2.3E-04
TMED10.6506.54.3	rs4589296	G	Men	No	-0.011 (-0.080,0.058)	7.5E-01	0.185 (0.119,0.252)	4.5E-08	5.8E-05
TMEM132C.11128.29.3	rs2501941	C	Women	No	0.285 (0.185,0.386)	2.7E-08	-0.042 (-0.143,0.060)	4.2E-01	7.3E-06
TMEM132C.7173.141.3	rs200508204	G	Men	No	-0.005 (-0.079,0.069)	8.9E-01	0.201 (0.128,0.273)	5.3E-08	9.5E-05
TMEM132D.13416.8.3	rs8064039	T	Women	Yes	-0.204 (-0.277,-0.131)	5.1E-08	0.093 (0.020,0.166)	1.2E-02	1.8E-08
TMEM132D.13416.8.3	rs73187233	C	Men	Yes	-0.045 (-0.135,0.045)	3.3E-01	0.250 (0.160,0.340)	4.4E-08	5.4E-06

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
TMEM230.11542.11.3	rs6082968	C	Men	No	0.066 (-0.007,0.140)	7.7E-02	-0.194 (-0.266,-0.123)	1.0E-07	6.5E-07
TMEM230.11542.11.3	rs56266868	G	Women	No	0.532 (0.351,0.714)	8.7E-09	-0.122 (-0.288,0.044)	1.5E-01	1.8E-07
TMEM57.12367.52.3	rs11023770	C	Women	Yes	-0.202 (-0.276,-0.128)	8.9E-08	0.044 (-0.031,0.120)	2.5E-01	4.9E-06
TMEM57.12367.52.3	rs12523973	T	Women	No	0.376 (0.243,0.509)	3.2E-08	-0.106 (-0.252,0.039)	1.5E-01	1.6E-06
TMEM59L.9959.60.3	rs852079	T	Men	No	0.007 (-0.073,0.087)	8.6E-01	-0.223 (-0.302,-0.144)	3.4E-08	6.1E-05
TMPRSS11B.10895.28.3	rs112352594	G	Women	No	0.568 (0.383,0.752)	1.7E-09	-0.039 (-0.247,0.170)	7.2E-01	1.9E-05
TMPRSS11D.6547.83.3	rs12904548	G	Women	No	-0.360 (-0.489,-0.231)	4.5E-08	-0.009 (-0.140,0.122)	9.0E-01	1.8E-04
TMPRSS5.8002.27.3	rs6037959	A	Men	No	-0.034 (-0.106,0.037)	3.5E-01	0.187 (0.119,0.254)	5.8E-08	1.0E-05
TMPRSS5.8002.27.3	rs140430972	C	Women	No	-0.546 (-0.743,-0.349)	5.4E-08	0.051 (-0.149,0.251)	6.2E-01	3.0E-05
TNF.5936.53.3	rs575184	G	Women	No	0.223 (0.146,0.301)	1.6E-08	-0.004 (-0.079,0.071)	9.1E-01	3.5E-05
TNF.5936.53.3	rs9265496	C	Women	No	0.218 (0.141,0.296)	2.9E-08	0.013 (-0.063,0.090)	7.3E-01	2.1E-04
TNFAIP6.5036.50.1	rs11771844	A	Women	No	-0.383 (-0.518,-0.249)	2.4E-08	0.062 (-0.063,0.188)	3.3E-01	2.1E-06
TNFRSF10B.7693.13.3	rs12456737	G	Men	No	-0.035 (-0.107,0.037)	3.4E-01	0.195 (0.125,0.265)	5.4E-08	7.7E-06
TNFRSF13B.2704.74.1	rs5829222	C	Men	Yes	-0.126 (-0.287,0.036)	1.3E-01	0.407 (0.258,0.556)	9.0E-08	2.1E-06
TNFRSF1A.2654.19.1	rs12498735	C	Men	No	0.033 (-0.038,0.104)	3.6E-01	-0.205 (-0.274,-0.136)	6.9E-09	2.7E-06
TNNI3.5441.67.3	rs74943542	A	Men	No	-0.011 (-0.085,0.063)	7.7E-01	0.198 (0.126,0.270)	7.8E-08	7.4E-05
TNS2.11667.29.3	rs6912572	T	Men	No	-0.006 (-0.085,0.073)	8.9E-01	-0.209 (-0.283,-0.134)	3.8E-08	2.5E-04
TNS4.9927.96.3	rs1548304	C	Men	No	0.009 (-0.062,0.081)	7.9E-01	-0.188 (-0.255,-0.121)	4.2E-08	8.1E-05
TOPBP1.9947.22.3	rs6821	C	Women	No	0.221 (0.143,0.298)	2.5E-08	0.011 (-0.066,0.088)	7.9E-01	1.7E-04
TP53.6152.111.3	rs2350913	G	Women	No	0.243 (0.162,0.325)	4.8E-09	-0.007 (-0.090,0.076)	8.6E-01	2.4E-05
TPSG1.14103.12.3	rs182162144	A	Women	No	0.447 (0.285,0.609)	6.9E-08	-0.016 (-0.166,0.134)	8.3E-01	4.0E-05
TPST1.7928.183.3	rs3759542	G	Men	No	0.009 (-0.141,0.160)	9.0E-01	-0.403 (-0.543,-0.263)	1.7E-08	8.5E-05
TPST2.8024.64.3	rs7734857	T	Women	Yes	0.217 (0.140,0.294)	3.3E-08	-0.021 (-0.097,0.054)	5.8E-01	1.4E-05
TREML1.9329.28.3	rs12509727	T	Men	Yes	-0.054 (-0.128,0.019)	1.5E-01	0.197 (0.126,0.267)	4.3E-08	1.4E-06
TSN.12477.42.3	rs2047422	A	Women	Yes	0.202 (0.130,0.274)	3.9E-08	0.003 (-0.067,0.072)	9.4E-01	9.8E-05

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
TSNARE1.7069.9.4	rs4849274	G	Women	No	0.218 (0.144,0.292)	7.0E-09	-0.003 (-0.074,0.069)	9.4E-01	2.5E-05
TSTA3.12657.2.3	rs62358068	A	Men	No	-0.020 (-0.102,0.061)	6.3E-01	0.217 (0.138,0.296)	6.8E-08	4.0E-05
TXLNA.9065.28.3	rs77131492	A	Women	No	-0.523 (-0.708,-0.338)	2.9E-08	-0.041 (-0.220,0.139)	6.6E-01	2.4E-04
TXN.8813.160.3	rs274472	T	Women	No	0.384 (0.255,0.513)	5.9E-09	0.031 (-0.095,0.156)	6.3E-01	1.2E-04
TYK2.5260.80.3	rs73166407	G	Women	No	0.239 (0.152,0.327)	8.4E-08	0.008 (-0.077,0.093)	8.6E-01	2.1E-04
TYRO3.2611.72.2	rs62124843	C	Men	No	-0.054 (-0.265,0.156)	6.1E-01	0.580 (0.367,0.793)	9.2E-08	3.2E-05
UBE2G2.9199.6.3	rs144562925	C	Women	No	-0.647 (-0.882,-0.412)	7.1E-08	0.164 (-0.086,0.413)	2.0E-01	3.7E-06
UBE2I.2877.3.1	rs11088998	G	Women	Yes	0.244 (0.159,0.329)	1.7E-08	-0.028 (-0.113,0.056)	5.1E-01	8.4E-06
UBXN4.9970.7.3	rs115910097	C	Men	No	0.064 (-0.115,0.243)	4.8E-01	-0.472 (-0.643,-0.301)	6.7E-08	2.2E-05
UBXN4.9997.12.3	rs12693374	A	Women	No	-0.243 (-0.330,-0.156)	4.0E-08	0.015 (-0.068,0.098)	7.3E-01	2.6E-05
UFM1.3836.51.2	rs11640887	C	Women	No	0.220 (0.143,0.298)	2.7E-08	-0.038 (-0.115,0.038)	3.3E-01	3.4E-06
UGT1A8.8899.75.3	rs9345317	T	Men	No	-0.045 (-0.131,0.041)	3.1E-01	0.230 (0.148,0.313)	4.8E-08	6.1E-06
UNC5A.7975.97.3	rs13088553	T	Women	No	-0.210 (-0.286,-0.135)	4.6E-08	0.004 (-0.070,0.078)	9.2E-01	7.5E-05
UROS.11248.43.3	rs73134208	C	Women	No	-0.332 (-0.445,-0.219)	8.2E-09	0.081 (-0.029,0.190)	1.5E-01	2.7E-07
USP25.9215.117.3	rs144858068	C	Women	Yes	0.421 (0.269,0.573)	5.3E-08	-0.070 (-0.232,0.092)	4.0E-01	1.4E-05
VAMP3.7903.18.3	rs7679436	C	Women	No	-0.228 (-0.311,-0.146)	6.3E-08	0.054 (-0.027,0.135)	1.9E-01	1.7E-06
VCL.8750.46.3	rs10863126	T	Men	No	0.008 (-0.066,0.082)	8.3E-01	-0.209 (-0.281,-0.137)	1.2E-08	3.9E-05
VTN.8280.238.3	rs117128433	C	Women	No	-0.558 (-0.763,-0.352)	1.0E-07	0.149 (-0.061,0.360)	1.6E-01	2.4E-06
VTN.8280.238.3	rs9342196	C	Women	No	0.246 (0.158,0.334)	3.8E-08	-0.017 (-0.102,0.067)	6.9E-01	2.3E-05
VWC2L.7995.16.3	rs12365739	C	Men	Yes	0.081 (-0.009,0.172)	7.8E-02	-0.244 (-0.333,-0.155)	7.8E-08	5.1E-07
WFDC5.6969.14.3	rs10089128	C	Men	No	-0.056 (-0.242,0.130)	5.5E-01	0.539 (0.346,0.731)	4.0E-08	1.3E-05
WNK3.5493.17.3	rs56726300	C	Women	Yes	-0.489 (-0.657,-0.322)	1.1E-08	0.058 (-0.100,0.215)	4.8E-01	3.3E-06
XDH.11264.33.3	rs113593110	C	Women	Yes	0.362 (0.233,0.491)	3.8E-08	-0.085 (-0.211,0.041)	1.9E-01	1.2E-06
XRCC4.9886.28.3	rs10975139	C	Women	No	0.196 (0.125,0.268)	6.4E-08	-0.014 (-0.082,0.055)	7.0E-01	3.2E-05

SOMAMER ID	rsID	EA	Index SNP	High confidence sexdiff	Women estimate (95% CI)	Women P_{GWAS}	Men estimate (95% CI)	Men P_{GWAS}	$P_{sexdiff}$
YWHAB.YWHAE.YWHAG.YWHAH.YWHAQ.YWHAZ.SFN.4707.50.2	rs10952071	T	Women	No	0.273 (0.173,0.374)	9.9E-08	-0.030 (-0.136,0.075)	5.8E-01	4.6E-05
ZAP70.3837.6.2	rs62402466	T	Women	No	-0.305 (-0.416,-0.195)	5.7E-08	-0.009 (-0.118,0.100)	8.7E-01	1.8E-04
ZFP91.13651.54.3	rs2505722	G	Men	No	-0.015 (-0.084,0.055)	6.8E-01	0.196 (0.127,0.265)	2.8E-08	2.6E-05
ZNF174.11486.26.3	rs71357272	C	Men	No	0.004 (-0.088,0.095)	9.4E-01	-0.251 (-0.343,-0.158)	1.0E-07	1.3E-04
ZNF23.11565.58.3	rs1788484	C	Women	No	0.205 (0.130,0.279)	7.2E-08	-0.036 (-0.110,0.038)	3.4E-01	7.2E-06
ZNF329.12803.9.3	rs7606391	A	Women	No	-0.205 (-0.276,-0.133)	2.3E-08	0.007 (-0.063,0.077)	8.5E-01	3.7E-05
ZNF75D.11596.47.3	rs2193400	C	Men	No	0.018 (-0.073,0.110)	6.9E-01	0.264 (0.172,0.355)	1.7E-08	2.1E-04

Only pQTLs that were associated with the protein P_{GWAS}<1% FDR, passed the quality control (consistent effect directions P_{GWAS}<0.05 in both INTERVAL cohorts), and with $P_{sexdiff}$<1% FDR are included in the table due to space restrictions (Benjamini and Hochberg, 1995). Index SNP denotes whether the SNP was an index pQTL in men, women, or both sexes. High confidence sexdiff denotes if the pQTL had consistent sex-differential effect estimates for all protein transformation and adjustment protocols, see section 3.2.9.4.2 for full criteria. Effect estimates as change in sex-specific SD units of the protein per additional EA.

Appendix Table 3.3. Mapping between aptamer IDs and full protein names for aptamers associated with obesity and cardiometabolic diseases in the MR analyses.

Exposure, SOMAMER ID	Exposure
ABO.9253.52.3	Histo-blood group ABO system transferase
ACP1.3858.5.1	Low molecular weight phosphotyrosine protein phosphatase
ADAMTS5.3168.8.2	A disintegrin and metalloproteinase with thrombospondin motifs 5
AGER.4125.52.2	Advanced glycosylation end product-specific receptor, soluble
AGRP.2813.11.2	Agouti-related protein
AKR1C1.12618.50.3	Aldo-keto reductase family 1 member C1
ASIP.5676.54.3	Agouti-signaling protein
ASPN.6451.64.3	Asporin
ATP1B2.7218.87.3	Sodium/potassium-transporting ATPase subunit beta-2
B3GAT3.6897.38.3	Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3
C1QTNF5.7810.20.3	Complement C1q tumor necrosis factor-related protein 5
CFH.4159.130.1	Complement factor H
CNTN2.3296.92.2	Contactin-2
COL15A1.8974.172.3	Collagen alpha-1(XV) chain
COLEC11.4430.44.3	Collectin-11
CPB1.6356.3.3	Carboxypeptidase B
CST7.3302.58.1	Cystatin-F
CTRB1.5671.1.3	Chymotrypsinogen B
CTSB.3061.61.2	Cathepsin B
DNAJC30.7866.11.3	DnaJ homolog subfamily C member 30
ERO1LB.7994.41.3	ERO1-like protein beta
FAM3D.13102.1.3	Protein FAM3D
FUT10.7156.2.3	Alpha-(1,3)-fucosyltransferase 10
GDF15.4374.45.2	Growth/differentiation factor 15
GRN.4992.49.1	Granulins
HLA.DQA2.7757.5.3	HLA class II histocompatibility antigen, DQ alpha 2 chain
HPGDS.12549.33.3	Hematopoietic prostaglandin D synthase
IGFBP7.3320.49.2	Insulin-like growth factor-binding protein 7
IL6R.4139.71.2	Interleukin-6 receptor subunit alpha
IL6R.8092.29.3	Interleukin-6 receptor subunit alpha
IL6ST.2620.4.2	Interleukin-6 receptor subunit beta
ITIH1.7955.195.3	Inter-alpha-trypsin inhibitor heavy chain H1
ITIH5.8233.2.3	Inter-alpha-trypsin inhibitor heavy chain H5
KYNU.4559.64.2	Kynureninase
LCT.9017.58.3	Lactase-phlorizin hydrolase
LYZ.4920.10.1	Lysozyme C
MAN2B2.9251.28.3	Epididymis-specific alpha-mannosidase
MANBA.6382.17.3	Beta-mannosidase
MANSC4.9578.263.3	MANSC domain-containing protein 4
MFAP2.9294.45.3	Microfibrillar-associated protein 2

Exposure, SOMAMER ID	Exposure
MICB.5102.55.3	MHC class I polypeptide-related sequence B
NCR3.3003.29.2	Natural cytotoxicity triggering receptor 3
NELL1.6544.33.3	Protein kinase C-binding protein NELL1
NPPB.7655.11.3	N-terminal pro-BNP
NPTX1.9256.78.3	Neuronal pentraxin-1
NQO1.9837.60.3	NAD(P)H dehydrogenase [quinone] 1
PAM.5620.13.3	Peptidyl-glycine alpha-amidating monooxygenase
PAM.5620.13.3	Peptidyl-glycine alpha-amidating monooxygenase
PCSK1.13388.57.3	Neuroendocrine convertase 1
PCSK7.4459.68.2	Proprotein convertase subtilisin/kexin type 7
QPCTL.8866.53.3	Glutaminyl-peptide cyclotransferase-like protein
RSPO3.13094.75.3	R-spondin-3
SCG3.7957.2.3	Secretogranin-3
SCG3.7957.2.3	Secretogranin-3
SEMA5A.13132.14.3	Semaphorin-5A
TAPBP.12378.71.3	Tapasin
TEK.3773.15.4	Angiopoietin-1 receptor, soluble
THSD1.5621.64.3	Thrombospondin type-1 domain-containing protein 1
TIMP4.6462.12.3	Metalloproteinase inhibitor 4
TNFSF12.5939.42.3	Tumor necrosis factor ligand superfamily member 12
WFIKKN2.13408.23.3	WAP, Kazal, immunoglobulin, Kunitz and NTR domain-containing protein 2
XXYLT1.6375.75.3	Xyloside xylosyltransferase 1

Appendix Table 3.4. Sex-specific effect estimates of proteins with effects on obesity.

Exposure, SOMAMER ID	Out-come	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W estimate (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M estimate (95% CI)	M P_{MR}	M MR tier
ABO.9253.52.3	WHR	Cis pQTLs	IVW	5.8E-05	low	2 (0)	-0.001 (-0.004, 0.003)	0.626	-	2 (0)	0.010 (0.006, 0.014)	7.4E-07	low
ABO.9253.52.3	WHR	Index pQTLs	IVW	5.8E-05	low	2 (0)	-0.001 (-0.004, 0.003)	0.626	-	2 (0)	0.010 (0.006, 0.014)	7.4E-07	low
ABO.9253.52.3	WHR	Women index pQTLs, men weights	IVW	5.8E-05	low	-	-	-	-	2 (0)	0.010 (0.006, 0.014)	7.4E-07	low
ABO.9253.52.3	WHRadj BMI	Cis pQTLs	IVW	5.5E-08	low	2 (0)	-0.003 (-0.007, 0.001)	0.106	-	2 (0)	0.012 (0.008, 0.016)	5.2E-09	low
ABO.9253.52.3	WHRadj BMI	Index pQTLs	IVW	5.5E-08	low	2 (0)	-0.003 (-0.007, 0.001)	0.106	-	2 (0)	0.012 (0.008, 0.016)	5.2E-09	low
ABO.9253.52.3	WHRadj BMI	Women index pQTLs, men weights	IVW	5.5E-08	low	-	-	-	-	2 (0)	0.012 (0.008, 0.016)	5.2E-09	low
ACP1.3858.5.1	WHRadj BMI	Cis pQTLs	IVW	0.003	-	1 (0)	0.001 (-0.004, 0.006)	0.617	-	1 (0)	0.012 (0.007, 0.018)	8.5E-06	high
ACP1.3858.5.1	WHRadj BMI	Index pQTLs	IVW	0.003	-	1 (0)	0.001 (-0.004, 0.006)	0.617	-	1 (0)	0.012 (0.007, 0.018)	8.5E-06	high
ACP1.3858.5.1	WHRadj BMI	Women index pQTLs, men weights	IVW	0.003	-	-	-	-	-	1 (0)	0.012 (0.007, 0.018)	8.5E-06	high
AGER.4125.52.2	BMI	Cis pQTLs	IVW	0.207	-	1 (0)	-0.006 (-0.024, 0.012)	0.516	-	1 (0)	-0.022 (-0.039, -0.005)	0.011	low

Exposure, SOMAMER ID	Out-come	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W estimate (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M estimate (95% CI)	M P_{MR}	M MR tier
AGER.4125.52.2	BMI	Index pQTLs	IVW	0.232	-	5 (0)	-0.015 (-0.026, -0.005)	0.005	-	5 (0)	-0.026 (-0.039, -0.012)	1.4E-04	low
AGER.4125.52.2	BMI	Index pQTLs	Simple median	0.346	-	5 (0)	-0.012 (-0.027, 0.003)	0.108	-	5 (0)	-0.022 (-0.037, -0.008)	0.003	low
AGER.4125.52.2	BMI	Index pQTLs	Weighted median	0.313	-	5 (0)	-0.012 (-0.026, 0.002)	0.081	-	5 (0)	-0.022 (-0.036, -0.008)	0.002	low
AGER.4125.52.2	BMI	Women index pQTLs, men weights	IVW	0.040	-	-	-	-	-	5 (0)	-0.035 (-0.050, -0.019)	1.1E-05	low
AGER.4125.52.2	BMI	Women index pQTLs, men weights	Simple median	0.124	-	-	-	-	-	5 (0)	-0.034 (-0.056, -0.011)	0.004	low
AGER.4125.52.2	BMI	Women index pQTLs, men weights	Weighted median	0.037	-	-	-	-	-	5 (0)	-0.037 (-0.056, -0.018)	1.0E-04	low
AKR1C1.12618.50.3	WHR	Cis pQTLs	IVW	1.9E-05	high	1 (1)	-0.007 (-0.026, 0.013)	0.505	-	1 (0)	-0.074 (-0.098, -0.050)	1.0E-09	high
AKR1C1.12618.50.3	WHR	Index pQTLs	IVW	1.9E-05	high	1 (1)	-0.007 (-0.026, 0.013)	0.505	-	1 (0)	-0.074 (-0.098, -0.050)	1.0E-09	high
AKR1C1.12618.50.3	WHR	Women index pQTLs, men weights	IVW	4.3E-06	high	-	-	-	-	1 (1)	-0.087 (-0.115, -0.059)	1.1E-09	high
AKR1C1.12618.50.3	WHRadj BMI	Cis pQTLs	IVW	2.7E-08	high	1 (1)	-0.007 (-0.027, 0.012)	0.470	-	1 (0)	-0.096 (-0.121, -0.072)	9.6E-15	high
AKR1C1.12618.50.3	WHRadj BMI	Index pQTLs	IVW	2.7E-08	high	1 (1)	-0.007 (-0.027, 0.012)	0.470	-	1 (0)	-0.096 (-0.121, -0.072)	9.6E-15	high

Exposure, SOMAMER ID	Out-come	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W estimate (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M estimate (95% CI)	M P_{MR}	M MR tier
AKR1C1.12618.5 0.3	WHRadj BMI	Women index pQTLs, men weights	IVW	1.9E-08	high	-	-	-	-	1 (1)	-0.105 (-0.133, -0.077)	1.4E-13	high
ASPN.6451.64.3	WHRadj BMI	Cis pQTLs	IVW	0.627	-	1 (0)	0.024 (0.014, 0.034)	6.2E-06	high	1 (0)	0.028 (0.016, 0.039)	3.4E-06	high
ASPN.6451.64.3	WHRadj BMI	Index pQTLs	IVW	0.627	-	1 (0)	0.024 (0.014, 0.034)	6.2E-06	high	1 (0)	0.028 (0.016, 0.039)	3.4E-06	high
ASPN.6451.64.3	WHRadj BMI	Women index pQTLs, men weights	IVW	0.651	-	-	-	-	-	1 (0)	0.028 (0.016, 0.039)	5.7E-06	high
CTSB.3061.61.2	WHRadj BMI	Cis pQTLs	IVW	0.012	-	1 (0)	0.021 (0.008, 0.034)	0.002	-	1 (0)	0.049 (0.032,0.06 6)	1.6E-08	high
CTSB.3061.61.2	WHRadj BMI	Index pQTLs	IVW	0.012	-	1 (0)	0.021 (0.008, 0.034)	0.002	-	1 (0)	0.049 (0.032, 0.066)	1.6E-08	high
CTSB.3061.61.2	WHRadj BMI	Women index pQTLs, men weights	IVW	0.013	-	-	-	-	-	1 (0)	0.049 (0.031, 0.066)	4.5E-08	high
FUT10.7156.2.3	BMI	Cis pQTLs	IVW	0.242	-	1 (0)	-0.017 (-0.028, -0.007)	0.001	-	1 (0)	-0.027 (-0.038, -0.016)	2.7E-06	high
FUT10.7156.2.3	BMI	Index pQTLs	IVW	0.242	-	1 (0)	-0.017 (-0.028, -0.007)	0.001	-	1 (0)	-0.027 (-0.038, -0.016)	2.7E-06	high
FUT10.7156.2.3	BMI	Women index pQTLs, men weights	IVW	0.161	-	-	-	-	-	1 (0)	-0.029 (-0.040, -0.017)	8.5E-07	high
FUT10.7156.2.3	WHR	Cis pQTLs	IVW	0.060	-	1 (0)	-0.012 (-0.023, -0.002)	0.022	-	1 (0)	-0.028 (-0.039, -0.016)	3.1E-06	high

Exposure, SOMAMER ID	Out-come	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W estimate (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M estimate (95% CI)	M P_{MR}	M MR tier
FUT10.7156.2.3	WHR	Index pQTLs	IVW	0.060	-	1 (0)	-0.012 (-0.023, -0.002)	0.022	-	1 (0)	-0.028 (-0.039, -0.016)	3.1E-06	high
FUT10.7156.2.3	WHR	Women index pQTLs, men weights	IVW	0.040	-	-	-	-	-	1 (0)	-0.030 (-0.042, -0.017)	2.4E-06	high
GDF15.4374.45.2	BMI	Cis pQTLs	IVW	0.501	-	1 (0)	0.017 (0.007, 0.027)	9.7E-04	-	1 (0)	0.022 (0.011, 0.033)	8.1E-05	high
GDF15.4374.45.2	BMI	Index pQTLs	IVW	0.501	-	1 (0)	0.017 (0.007, 0.027)	9.7E-04	-	1 (0)	0.022 (0.011, 0.033)	8.1E-05	high
GDF15.4374.45.2	BMI	Women index pQTLs, men weights	IVW	0.428	-	-	-	-	-	1 (0)	0.023 (0.012, 0.034)	3.8E-05	high
HLA.DQA2.7757.5.3	WHRadj BMI	Cis pQTLs	IVW	0.547	-	3 (3)	0.041 (0.032, 0.050)	2.8E-18	-	2 (2)	0.046 (0.033, 0.059)	1.4E-11	low
HLA.DQA2.7757.5.3	WHRadj BMI	Index pQTLs	IVW	0.700	-	4 (3)	0.040 (0.016, 0.065)	0.001	-	2 (2)	0.046 (0.033, 0.059)	1.4E-11	low
HLA.DQA2.7757.5.3	WHRadj BMI	Women index pQTLs, men weights	IVW	0.903	-	-	-	-	-	4 (3)	0.043 (0.011, 0.075)	0.008	low
HLA.DQA2.7757.5.3	WHRadj BMI	Women index pQTLs, men weights	Simple median	0.538	-	-	-	-	-	4 (3)	0.048 (0.014, 0.081)	0.005	low
HLA.DQA2.7757.5.3	WHRadj BMI	Women index pQTLs, men weights	Weighted median	0.739	-	-	-	-	-	4 (3)	0.049 (0.034, 0.065)	1.1E-09	low
HPGDS.12549.33.3	WHR	Cis pQTLs	IVW	0.001	high	1 (0)	0.000 (-0.012, 0.012)	1.000	-	1 (0)	0.029 (0.016, 0.042)	1.3E-05	high

Exposure, SOMAMER ID	Out-come	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W estimate (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M estimate (95% CI)	M P_{MR}	M MR tier
HPGDS.12549.33.3	WHR	Index pQTLs	IVW	0.001	high	1 (0)	0.000 (-0.012,0.012)	1.000	-	1 (0)	0.029 (0.016, 0.042)	1.3E-05	high
HPGDS.12549.33.3	WHR	Women index pQTLs, men weights	IVW	0.002	high	-	-	-	-	1 (0)	0.029 (0.016, 0.042)	2.0E-05	high
IL6ST.2620.4.2	WHR	Cis pQTLs	IVW	0.582	-	1 (0)	-0.014 (-0.036, 0.008)	0.207	-	1 (1)	-0.022 (-0.041, -0.004)	0.018	low
IL6ST.2620.4.2	WHR	Index pQTLs	IVW	0.013	-	2 (0)	-0.003 (-0.020, 0.014)	0.754	-	2 (1)	-0.032 (-0.047, -0.016)	5.4E-05	low
IL6ST.2620.4.2	WHR	Women index pQTLs, men weights	IVW	0.725	-	-	-	-	-	2 (0)	0.002 (-0.018, 0.022)	0.844	low
IL6ST.2620.4.2	WHRadj BMI	Cis pQTLs	IVW	0.373	-	1 (0)	-0.014 (-0.036, 0.008)	0.216	-	1 (1)	-0.027 (-0.046, -0.009)	0.004	low
IL6ST.2620.4.2	WHRadj BMI	Index pQTLs	IVW	0.008	-	2 (0)	-0.006 (-0.023, 0.011)	0.491	-	2 (1)	-0.037 (-0.052, -0.022)	2.0E-06	low
IL6ST.2620.4.2	WHRadj BMI	Women index pQTLs, men weights	IVW	0.755	-	-	-	-	-	2 (0)	-0.010 (-0.031, 0.011)	0.335	low
LCT.9017.58.3	BMI	Cis pQTLs	IVW	0.745	-	1 (1)	0.015 (0.009, 0.021)	9.0E-07	high	1 (1)	0.013 (0.007, 0.020)	7.4E-05	high
LCT.9017.58.3	BMI	Index pQTLs	IVW	0.723	-	1 (1)	0.015 (0.009, 0.021)	9.0E-07	high	2 (1)	0.013 (0.007, 0.019)	5.5E-05	high

Exposure, SOMAMER ID	Out-come	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W estimate (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M estimate (95% CI)	M P_{MR}	M MR tier
LCT.9017.58.3	BMI	Women index pQTLs, men weights	IVW	0.655	-	-	-	-	-	1 (1)	0.017 (0.010, 0.023)	3.1E-07	high
MANBA.6382.17.3	BMI	Cis pQTLs	IVW	0.402	-	1 (0)	-0.015 (-0.024, -0.007)	4.7E-04	-	1 (0)	-0.021 (-0.029, -0.012)	4.7E-06	high
MANBA.6382.17.3	BMI	Index pQTLs	IVW	0.402	-	1 (0)	-0.015 (-0.024, -0.007)	4.7E-04	-	1 (0)	-0.021 (-0.029, -0.012)	4.7E-06	high
MANBA.6382.17.3	BMI	Women index pQTLs, men weights	IVW	0.357	-	-	-	-	-	1 (0)	-0.021 (-0.031, -0.012)	6.8E-06	high
MFAP2.9294.45.3	BMI	Cis pQTLs	IVW	0.813	-	1 (0)	-0.040 (-0.066, -0.013)	0.004	-	1 (0)	-0.044 (-0.066,-0.022)	1.0E-04	high
MFAP2.9294.45.3	BMI	Index pQTLs	IVW	0.003	-	2 (0)	-0.002 (-0.019, 0.015)	0.811	-	1 (0)	-0.044 (-0.066, -0.022)	1.0E-04	high
MFAP2.9294.45.3	BMI	Women index pQTLs, men weights	IVW	0.963	-	-	-	-	-	2 (0)	-0.003 (-0.019, 0.014)	0.754	high
MICB.5102.55.3	WHR	Cis pQTLs	IVW	0.004	-	3 (0)	0.014 (0.006, 0.023)	6.3E-04	-	3 (1)	0.034 (0.023, 0.044)	9.4E-11	low
MICB.5102.55.3	WHR	Cis pQTLs	Simple median	0.039	-	3 (0)	0.003 (-0.014, 0.021)	0.704	-	3 (1)	0.034 (0.011, 0.057)	0.004	low
MICB.5102.55.3	WHR	Cis pQTLs	Weighted median	0.007	-	3 (0)	0.010 (-0.002, 0.022)	0.111	-	3 (1)	0.032 (0.021, 0.043)	1.5E-08	low
MICB.5102.55.3	WHR	Index pQTLs	IVW	0.109	-	4 (0)	0.016 (-0.001, 0.033)	0.065	-	4 (1)	0.032 (0.022, 0.041)	6.2E-11	low

Exposure, SOMAMER ID	Out-come	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W estimate (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M estimate (95% CI)	M P_{MR}	M MR tier
MICB.5102.55.3	WHR	Index pQTLs	Simple median	0.130	-	4 (0)	0.015 (0.003, 0.028)	0.013	-	4 (1)	0.033 (0.014, 0.052)	7.0E-04	low
MICB.5102.55.3	WHR	Index pQTLs	Weighted median	0.619	-	4 (0)	0.028 (0.014, 0.041)	5.2E-05	-	4 (1)	0.032 (0.021, 0.043)	1.8E-08	low
MICB.5102.55.3	WHR	Women index pQTLs, men weights	IVW	0.503	-	-	-	-	-	4 (0)	0.023 (0.010, 0.037)	8.3E-04	low
MICB.5102.55.3	WHR	Women index pQTLs, men weights	Simple median	0.979	-	-	-	-	-	4 (0)	0.016 (0.001, 0.030)	0.032	low
MICB.5102.55.3	WHR	Women index pQTLs, men weights	Weighted median	0.677	-	-	-	-	-	4 (0)	0.024 (0.013, 0.035)	1.8E-05	low
MICB.5102.55.3	WHRadj BMI	Cis pQTLs	IVW	1.5E-07	-	3 (0)	0.014 (0.006, 0.022)	9.9E-04	-	3 (1)	0.049 (0.039, 0.059)	4.7E-21	low
MICB.5102.55.3	WHRadj BMI	Cis pQTLs	Simple median	0.029	-	3 (0)	-0.001 (-0.022, 0.021)	0.959	-	3 (1)	0.036 (0.011, 0.061)	0.004	low
MICB.5102.55.3	WHRadj BMI	Cis pQTLs	Weighted median	6.6E-05	-	3 (0)	0.010 (-0.006, 0.025)	0.229	-	3 (1)	0.049 (0.037, 0.061)	4.9E-16	low
MICB.5102.55.3	WHRadj BMI	Index pQTLs	IVW	0.065	-	4 (0)	0.016 (-0.015, 0.047)	0.304	-	4 (1)	0.047 (0.036, 0.058)	5.2E-17	low
MICB.5102.55.3	WHRadj BMI	Index pQTLs	Simple median	0.124	-	4 (0)	0.017 (0.003, 0.032)	0.020	-	4 (1)	0.036 (0.017, 0.054)	1.8E-04	low
MICB.5102.55.3	WHRadj BMI	Index pQTLs	Weighted median	0.297	-	4 (0)	0.035 (0.016, 0.054)	3.3E-04	-	4 (1)	0.047 (0.035, 0.059)	2.9E-15	low

Exposure, SOMAMER ID	Out-come	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W estimate (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M estimate (95% CI)	M P_{MR}	M MR tier
MICB.5102.55.3	WHRadj BMI	Women index pQTLs, men weights	IVW	0.454	-	-	-	-	-	4 (0)	0.032 (0.004, 0.060)	0.025	low
MICB.5102.55.3	WHRadj BMI	Women index pQTLs, men weights	Simple median	0.879	-	-	-	-	-	4 (0)	0.019 (0.001, 0.037)	0.040	low
MICB.5102.55.3	WHRadj BMI	Women index pQTLs, men weights	Weighted median	0.578	-	-	-	-	-	4 (0)	0.042 (0.026, 0.059)	4.7E-07	low
NPTX1.9256.78.3	BMI	Cis pQTLs	IVW	0.349	-	1 (0)	-0.040 (-0.058, -0.022)	1.7E-05	high	1 (0)	-0.052 (-0.071, -0.034)	2.0E-08	high
NPTX1.9256.78.3	BMI	Index pQTLs	IVW	0.349	-	1 (0)	-0.040 (-0.058,-0.022)	1.7E-05	high	1 (0)	-0.052 (-0.071, -0.034)	2.0E-08	high
NPTX1.9256.78.3	BMI	Women index pQTLs, men weights	IVW	0.335	-	-	-	-	-	1 (0)	-0.053 (-0.072, -0.034)	3.2E-08	high
PAM.5620.13.3	BMI	Cis pQTLs	IVW	0.069	-	2 (0)	-0.009 (-0.017, 0.000)	0.057	-	2 (0)	-0.021 (-0.030, -0.011)	2.9E-05	high
PAM.5620.13.3	BMI	Index pQTLs	IVW	0.079	-	3 (0)	-0.009 (-0.017, -0.001)	0.027	-	2 (0)	-0.021 (-0.030, -0.011)	2.9E-05	high
PAM.5620.13.3	BMI	Women index pQTLs, men weights	IVW	0.568	-	-	-	-	-	3 (0)	-0.006 (-0.014, 0.003)	0.190	high
PAM.5620.13.3	BMI	Women index pQTLs, men weights	Simple median	0.134	-	-	-	-	-	3 (0)	-0.000 (-0.013, 0.012)	0.951	high
PAM.5620.13.3	BMI	Women index pQTLs, men weights	Weighted median	0.230	-	-	-	-	-	3 (0)	-0.006 (-0.016, 0.005)	0.294	high

Exposure, SOMAMER ID	Out-come	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W estimate (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M estimate (95% CI)	M P_{MR}	M MR tier
PAM.5620.13.3	WHRadj BMI	Cis pQTLs	IVW	0.348	-	2 (0)	0.015 (0.006, 0.024)	8.6E-04	low	2 (0)	0.021 (0.011, 0.032)	5.1E-05	high
PAM.5620.13.3	WHRadj BMI	Index pQTLs	IVW	0.408	-	3 (0)	0.016 (0.008, 0.024)	1.7E-04	low	2 (0)	0.021 (0.011, 0.032)	5.1E-05	high
PAM.5620.13.3	WHRadj BMI	Women index pQTLs, men weights	IVW	0.104	-	-	-	-	-	3 (0)	0.006 (-0.003, 0.015)	0.201	high
PAM.5620.13.3	WHRadj BMI	Women index pQTLs, men weights	Simple median	0.118	-	-	-	-	-	3 (0)	0.007 (-0.004, 0.018)	0.232	high
PAM.5620.13.3	WHRadj BMI	Women index pQTLs, men weights	Weighted median	0.042	-	-	-	-	-	3 (0)	0.005 (-0.005, 0.015)	0.377	high
PCSK1.13388.57.3	BMI	Cis pQTLs	IVW	0.106	-	1 (0)	-0.013 (-0.018, -0.009)	6.3E-09	high	1 (0)	-0.019 (-0.024, -0.014)	3.2E-13	high
PCSK1.13388.57.3	BMI	Index pQTLs	IVW	0.127	-	2 (0)	-0.014 (-0.018, -0.009)	1.4E-09	high	1 (0)	-0.019 (-0.024, -0.014)	3.2E-13	high
PCSK1.13388.57.3	BMI	Women index pQTLs, men weights	IVW	0.108	-	-	-	-	-	2 (0)	-0.019 (-0.025, -0.014)	1.5E-13	high
PCSK1.13388.57.3	WHR	Cis pQTLs	IVW	0.018	-	1 (0)	-0.003 (-0.008, 0.002)	0.182	-	1 (0)	-0.012 (-0.018, -0.007)	2.0E-05	high
PCSK1.13388.57.3	WHR	Index pQTLs	IVW	0.017	-	2 (0)	-0.003 (-0.008, 0.001)	0.168	-	1 (0)	-0.012 (-0.018, -0.007)	2.0E-05	high
PCSK1.13388.57.3	WHR	Women index pQTLs, men weights	IVW	0.018	-	-	-	-	-	2 (0)	-0.012 (-0.018, -0.006)	2.1E-05	high

Exposure, SOMAMER ID	Out-come	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W estimate (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M estimate (95% CI)	M P_{MR}	M MR tier
RSPO3.13094.75.3	WHR	Cis pQTLs	IVW	3.3E-13	high	1 (0)	0.158 (0.143, 0.174)	1.7E-90	high	1 (0)	0.068 (0.049, 0.087)	1.4E-12	high
RSPO3.13094.75.3	WHR	Index pQTLs	IVW	2.5E-05	high	2 (0)	0.117 (0.104, 0.131)	1.8E-68	high	1 (0)	0.068 (0.049, 0.087)	1.4E-12	high
RSPO3.13094.75.3	WHR	Women index pQTLs, men weights	IVW	0.004	high	-	-	-	-	2 (0)	0.079 (0.057, 0.102)	4.2E-12	high
RSPO3.13094.75.3	WHRadj BMI	Cis pQTLs	IVW	4.2E-15	high	1 (0)	0.182 (0.166, 0.197)	8.3E-119	high	1 (0)	0.084 (0.066, 0.103)	1.4E-18	high
RSPO3.13094.75.3	WHRadj BMI	Index pQTLs	IVW	9.5E-06	high	2 (0)	0.136 (0.123, 0.149)	9.7E-92	high	1 (0)	0.084 (0.066, 0.103)	1.4E-18	high
RSPO3.13094.75.3	WHRadj BMI	Women index pQTLs, men weights	IVW	0.002	high	-	-	-	-	2 (0)	0.094 (0.070, 0.117)	2.9E-15	high
SEMA5A.13132.14.3	WHRadj BMI	Cis pQTLs	IVW	0.125	-	1 (0)	-0.007 (-0.013, -0.000)	0.037	-	1 (0)	-0.014 (-0.021, -0.007)	1.3E-04	high
SEMA5A.13132.14.3	WHRadj BMI	Index pQTLs	IVW	0.125	-	1 (0)	-0.007 (-0.013, -0.000)	0.037	-	1 (0)	-0.014 (-0.021, -0.007)	1.3E-04	high
SEMA5A.13132.14.3	WHRadj BMI	Women index pQTLs, men weights	IVW	0.125	-	-	-	-	-	1 (0)	-0.014 (-0.021, -0.007)	1.3E-04	high
TAPBP.12378.71.3	BMI	Cis pQTLs	IVW	0.069	-	2 (0)	-0.008 (-0.019, 0.003)	0.157	-	2 (1)	-0.023 (-0.034, -0.011)	1.0E-04	high
TAPBP.12378.71.3	BMI	Index pQTLs	IVW	0.021	-	3 (0)	-0.005 (-0.015, 0.005)	0.344	-	2 (1)	-0.023 (-0.034, -0.011)	1.0E-04	high

Exposure, SOMAMER ID	Out-come	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W estimate (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M estimate (95% CI)	M P_{MR}	M MR tier
TAPBP.12378.71.3	BMI	Women index pQTLs, men weights	IVW	0.041	-	-	-	-	-	3 (0)	-0.021 (-0.033, -0.009)	6.0E-04	high
TAPBP.12378.71.3	BMI	Women index pQTLs, men weights	Simple median	0.233	-	-	-	-	-	3 (0)	-0.025 (-0.052, 0.002)	0.066	high
TAPBP.12378.71.3	BMI	Women index pQTLs, men weights	Weighted median	0.118	-	-	-	-	-	3 (0)	-0.021 (-0.034, -0.008)	0.001	high
TAPBP.12378.71.3	WHR	Cis pQTLs	IVW	0.033	-	2 (0)	-0.012 (-0.023, -0.001)	0.029	-	2 (1)	-0.030 (-0.041, -0.018)	4.4E-07	high
TAPBP.12378.71.3	WHR	Index pQTLs	IVW	0.021	-	3 (0)	-0.012 (-0.022, -0.002)	0.023	-	2 (1)	-0.030 (-0.041, -0.018)	4.4E-07	high
TAPBP.12378.71.3	WHR	Women index pQTLs, men weights	IVW	0.022	-	-	-	-	-	3 (0)	-0.030 (-0.043, -0.018)	1.2E-06	high
TAPBP.12378.71.3	WHR	Women index pQTLs, men weights	Simple median	0.254	-	-	-	-	-	3 (0)	-0.039 (-0.085, 0.006)	0.091	high
TAPBP.12378.71.3	WHR	Women index pQTLs, men weights	Weighted median	0.048	-	-	-	-	-	3 (0)	-0.030 (-0.044, -0.016)	2.4E-05	high
TAPBP.12378.71.3	WHRadj BMI	Cis pQTLs	IVW	0.184	-	2 (0)	-0.011 (-0.023, -0.000)	0.043	-	2 (1)	-0.022 (-0.034, -0.011)	1.5E-04	high
TAPBP.12378.71.3	WHRadj BMI	Index pQTLs	IVW	0.193	-	3 (0)	-0.012 (-0.022, -0.002)	0.019	-	2 (1)	-0.022 (-0.034, -0.011)	1.5E-04	high
TAPBP.12378.71.3	WHRadj BMI	Women index pQTLs, men weights	IVW	0.142	-	-	-	-	-	3 (0)	-0.024 (-0.036, -0.012)	1.2E-04	high

Exposure, SOMAMER ID	Out-come	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W estimate (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M estimate (95% CI)	M P_{MR}	M MR tier
TAPBP.12378.71.3	WHRadjBMI	Women index pQTLs, men weights	Simple median	0.388	-	-	-	-	-	3 (0)	-0.033 (-0.079, 0.013)	0.158	high
TAPBP.12378.71.3	WHRadjBMI	Women index pQTLs, men weights	Weighted median	0.180	-	-	-	-	-	3 (0)	-0.024 (-0.038, -0.010)	6.8E-04	high
TEK.3773.15.4	WHRadjBMI	Cis pQTLs	IVW	0.092	low	1 (0)	-0.014 (-0.034, 0.006)	0.177	-	1 (1)	0.015 (-0.012, 0.041)	0.277	low
TEK.3773.15.4	WHRadjBMI	Index pQTLs	IVW	0.001	low	3 (1)	-0.002 (-0.014, 0.010)	0.723	-	2 (1)	0.029 (0.014,0.044)	1.1E-04	low
TEK.3773.15.4	WHRadjBMI	Women index pQTLs, men weights	IVW	0.004	low	-	-	-	-	3 (1)	0.025 (0.011, 0.039)	4.4E-04	low
TEK.3773.15.4	WHRadjBMI	Women index pQTLs, men weights	Simple median	0.008	low	-	-	-	-	3 (1)	0.028 (0.004, 0.053)	0.025	low
TEK.3773.15.4	WHRadjBMI	Women index pQTLs, men weights	Weighted median	0.006	low	-	-	-	-	3 (1)	0.025 (0.010, 0.040)	0.001	low

Only proteins with an effect P_{MR}<1% FDR in the main analysis (using the index instrument and the IVW MR method) are included in the table due to space restrictions (Benjamini and Hochberg, 1995). MR tier denotes if the protein-obesity trait association is supported in the sensitivity analyes and P_{het} tier denotes if the protein-obesity trait association have robust sexually heterogeneous effect estimates in the sensitivity analyses, as outlined in section 3.2.8.2. Effect estimates as change in sex-stratified SD units of the obesity trait per 1 unit higher sex-stratified SD of the protein. "-" denotes that the analysis was not performed, either because of the how the methods pipeline was created or because the protein-outcome combination did not surpass the P_{MR}<1% FDR threshold in that sex or P_{het}<1% FDR in the sex-comparison (Benjamini and Hochberg, 1995). Prx, number of proxies; M, men; W, women.

Appendix Table 3.5. Sex-specific effect estimates of proteins with effects on cardiometabolic disease.

Exposure, SOMAMER ID	Outcome	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W OR (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M OR (95% CI)	M P_{MR}	M MR tier
CPB1.6356.3.3	T2D, all cases	Index pQTLs	IVW	0.004	-	1 (0)	0.91 (0.78, 1.05)	0.199	-	3 (0)	1.16 (1.08, 1.26)	1.1E-04	low
CPB1.6356.3.3	T2D, all cases	Index pQTLs	Simple median	-	-	-	-	-	-	3 (0)	1.01 (0.87, 1.17)	0.855	low
CPB1.6356.3.3	T2D, all cases	Index pQTLs	Weighted median	-	-	-	-	-	-	3 (0)	1.10 (0.98, 1.23)	0.116	low
CPB1.6356.3.3	T2D, all cases	Cis pQTLs	IVW	0.831	-	1 (0)	0.91 (0.78, 1.05)	0.199	-	1 (0)	0.93 (0.81, 1.06)	0.282	low
CPB1.6356.3.3	T2D, all cases	Women index pQTLs, men weights	IVW	0.846	-	-	-	-	-	1 (0)	0.92 (0.80, 1.07)	0.282	low
CTRB1.5671.1.3	T2D, all cases	Index pQTLs	IVW	0.001	low	1 (0)	0.97 (0.91, 1.03)	0.275	-	1 (0)	1.10 (1.05, 1.15)	4.7E-05	high
CTRB1.5671.1.3	T2D, all cases	Cis pQTLs	IVW	0.001	low	1 (0)	0.97 (0.91, 1.03)	0.275	-	1 (0)	1.10 (1.05, 1.15)	4.7E-05	high
CTRB1.5671.1.3	T2D, all cases	Women index pQTLs, men weights	IVW	0.378	low	-	-	-	-	1 (0)	1.00 (0.95, 1.05)	0.984	high
COLEC11.4430.44.3	CAD	Index pQTLs	IVW	0.161	-	3 (0)	1.02 (0.98, 1.06)	0.272	-	3 (0)	1.05 (1.03, 1.07)	2.4E-05	low
COLEC11.4430.44.3	CAD	Index pQTLs	Simple median	0.060	-	3 (0)	1.03 (0.97, 1.09)	0.382	-	3 (0)	1.12 (1.05, 1.19)	8.4E-04	low

Exposure, SOMAMER ID	Outcome	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W OR (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M OR (95% CI)	M P_{MR}	M MR tier
COLEC11.4430.44.3	CAD	Index pQTLs	Weighted median	0.147	-	3 (0)	1.02 (0.98, 1.06)	0.325	-	3 (0)	1.05 (1.03, 1.08)	4.1E-05	low
COLEC11.4430.44.3	CAD	Cis pQTLs	IVW	0.161	-	3 (0)	1.02 (0.98, 1.06)	0.272	-	3 (0)	1.05 (1.03, 1.07)	2.4E-05	low
COLEC11.4430.44.3	CAD	Cis pQTLs	Simple median	0.060	-	3 (0)	1.03 (0.97, 1.09)	0.382	-	3 (0)	1.12 (1.05, 1.19)	8.4E-04	low
COLEC11.4430.44.3	CAD	Cis pQTLs	Weighted median	0.147	-	3 (0)	1.02 (0.98, 1.06)	0.325	-	3 (0)	1.05 (1.03, 1.08)	4.1E-05	low
COLEC11.4430.44.3	CAD	Women index pQTLs, men weights	IVW	0.720	-	-	-	-	-	3 (0)	1.03 (0.98, 1.08)	0.193	low
COLEC11.4430.44.3	CAD	Women index pQTLs, men weights	Simple median	0.044	-	-	-	-	-	3 (0)	1.13 (1.05, 1.22)	9.9E-04	low
COLEC11.4430.44.3	CAD	Women index pQTLs, men weights	Weighted median	0.061	-	-	-	-	-	3 (0)	1.09 (1.02, 1.16)	0.006	low
ERO1LB.7994.41.3	T2D, all cases	Index pQTLs	IVW	0.100	-	2 (0)	1.07 (0.98, 1.16)	0.159	-	2 (0)	1.17 (1.10, 1.24)	1.0E-06	low
ERO1LB.7994.41.3	T2D, all cases	Cis pQTLs	IVW	0.222	-	1 (0)	1.12 (1.01, 1.23)	0.032	-	1 (0)	1.20 (1.12, 1.29)	1.7E-07	low
ERO1LB.7994.41.3	T2D, all cases	Women index pQTLs, men weights	IVW	0.020	-	-	-	-	-	2 (0)	1.22 (1.13, 1.32)	2.5E-07	low
GRN.4992.49.1	CAD	Index pQTLs	IVW	0.339	-	4 (0)	1.08 (1.02, 1.15)	0.007	-	3 (0)	1.12 (1.09, 1.15)	9.4E-16	low

Exposure, SOMAMER ID	Outcome	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W OR (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M OR (95% CI)	M P_{MR}	M MR tier
GRN.4992.49.1	CAD	Index pQTLs	Simple median	0.704	-	4 (0)	1.02 (0.96, 1.10)	0.469	-	3 (0)	1.05 (0.96, 1.14)	0.293	low
GRN.4992.49.1	CAD	Index pQTLs	Weighted median	0.185	-	4 (0)	1.09 (1.04, 1.13)	4.9E-05	-	3 (0)	1.13 (1.09, 1.17)	1.6E-11	low
GRN.4992.49.1	CAD	Cis pQTLs	IVW	0.857	-	1 (0)	0.96 (0.83, 1.11)	0.540	-	1 (0)	0.97 (0.91, 1.04)	0.381	low
GRN.4992.49.1	CAD	Women index pQTLs, men weights	IVW	0.747	-	-	-	-	-	4 (0)	1.13 (0.89, 1.42)	0.313	low
GRN.4992.49.1	CAD	Women index pQTLs, men weights	Simple median	0.042	-	-	-	-	-	4 (0)	1.18 (1.05, 1.34)	0.007	low
GRN.4992.49.1	CAD	Women index pQTLs, men weights	Weighted median	0.166	-	-	-	-	-	4 (0)	1.03 (0.96, 1.10)	0.476	low
IGFBP7.3320.49.2	T2D, all cases	Index pQTLs	IVW	0.005	-	1 (0)	1.05 (0.93, 1.17)	0.442	-	2 (1)	0.86 (0.79, 0.93)	8.3E-05	low
IGFBP7.3320.49.2	T2D, all cases	Cis pQTLs	IVW	0.063	-	1 (0)	1.05 (0.93, 1.17)	0.442	-	1 (0)	0.91 (0.83, 1.00)	0.041	low
IGFBP7.3320.49.2	T2D, all cases	Women index pQTLs, men weights	IVW	0.081	-	-	-	-	-	1 (0)	0.92 (0.84, 1.01)	0.066	low
IL6R.4139.71.2	CAD	Index pQTLs	IVW	0.515	-	2 (0)	0.97 (0.95, 1.00)	0.052	-	2 (0)	0.96 (0.94, 0.98)	1.2E-04	high
IL6R.8092.29.3	CAD	Index pQTLs	IVW	0.499	-	2 (0)	0.97 (0.95, 1.00)	0.057	-	2 (0)	0.96 (0.94, 0.98)	1.2E-04	high

Exposure, SOMAMER ID	Outcome	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W OR (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M OR (95% CI)	M P_{MR}	M MR tier
IL6R.4139.71.2	CAD	Cis pQTLs	IVW	0.515	-	2 (0)	0.97 (0.95, 1.00)	0.052	-	2 (0)	0.96 (0.94, 0.98)	1.2E-04	high
IL6R.8092.29.3	CAD	Cis pQTLs	IVW	0.499	-	2 (0)	0.97 (0.95, 1.00)	0.057	-	2 (0)	0.96 (0.94, 0.98)	1.2E-04	high
IL6R.4139.71.2	CAD	Women index pQTLs, men weights	IVW	0.517	-	-	-	-	-	2 (0)	0.96 (0.94, 0.98)	1.4E-04	high
IL6R.8092.29.3	CAD	Women index pQTLs, men weights	IVW	0.510	-	-	-	-	-	2 (0)	0.96 (0.94, 0.98)	1.2E-04	high
MANSC4.9578.263.3	T2D, all cases	Index pQTLs	IVW	0.429	-	1 (0)	0.88 (0.81, 0.95)	7.2E-04	-	1 (0)	0.91 (0.87, 0.95)	1.3E-04	high
MANSC4.9578.263.3	T2D, all cases	Cis pQTLs	IVW	0.429	-	1 (0)	0.88 (0.81, 0.95)	7.2E-04	-	1 (0)	0.91 (0.87, 0.95)	1.3E-04	high
MANSC4.9578.263.3	T2D, all cases	Women index pQTLs, men weights	IVW	0.449	-	-	-	-	-	1 (0)	0.91 (0.87, 0.95)	1.0E-04	high
PAM.5620.13.3	T2D, all cases	Index pQTLs	IVW	0.284	-	3 (0)	0.94 (0.89, 0.99)	0.032	-	2 (0)	0.90 (0.86, 0.95)	1.6E-04	high
PAM.5620.13.3	T2D, all cases	Cis pQTLs	IVW	0.497	-	2 (0)	0.93 (0.88, 0.98)	0.012	-	2 (0)	0.90 (0.86, 0.95)	1.6E-04	high
PAM.5620.13.3	T2D, all cases	Women index pQTLs, men weights	IVW	0.054	-	-	-	-	-	3 (0)	0.88 (0.85, 0.92)	6.5E-10	high
PAM.5620.13.3	T2D, all cases	Women index pQTLs, men weights	Simple median	0.227	-	-	-	-	-	3 (0)	0.92 (0.85, 1.00)	0.050	high

Exposure, SOMAMER ID	Outcome	Instrument type	Method	P_{het}	P_{het} tier	W N SNPs (prx)	W OR (95% CI)	W P_{MR}	W MR tier	M N SNPs (prx)	M OR (95% CI)	M P_{MR}	M MR tier
PAM.5620.13.3	T2D, all cases	Women index pQTLs, men weights	Weighted median	0.325	-	-	-	-	-	3 (0)	0.90 (0.86, 0.95)	1.8E-04	high
THSD1.5621.64.3	CAD	Index pQTLs	IVW	0.001	low	1 (0)	1.08 (0.99, 1.19)	0.094	-	2 (0)	0.91 (0.86, 0.95)	1.9E-04	low
THSD1.5621.64.3	CAD	Cis pQTLs	IVW	0.008	low	1 (0)	1.08 (0.99, 1.19)	0.094	-	1 (0)	0.92 (0.86, 0.99)	0.029	low
THSD1.5621.64.3	CAD	Women index pQTLs, men weights	IVW	0.008	low	-	-	-	-	1 (0)	0.92 (0.86, 0.99)	0.029	low

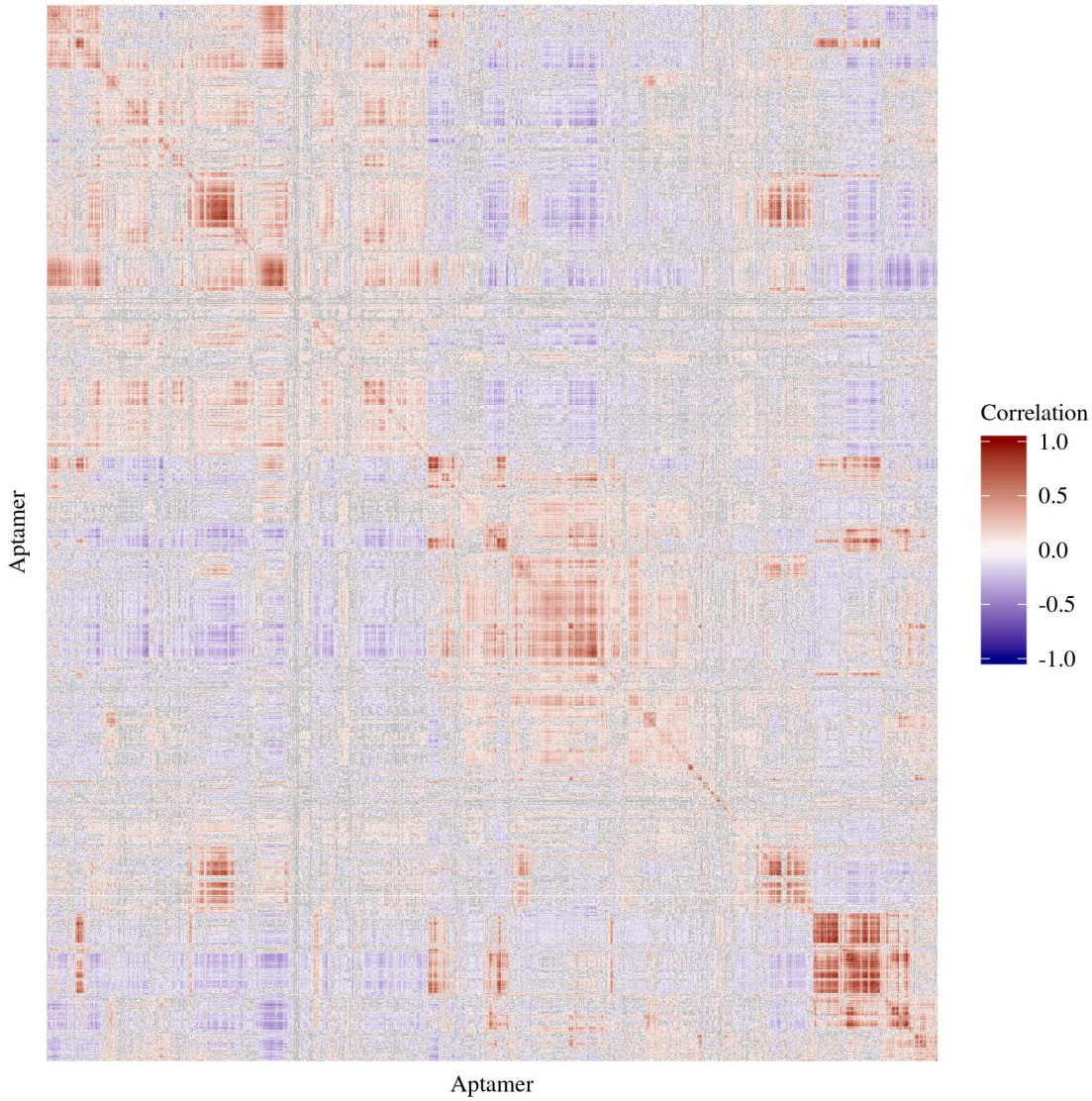
Only proteins with an effect P_{MR}<1% FDR in the main analysis (using the index instrument and the IVW MR method) are included in the table due to space restrictions (Benjamini and Hochberg, 1995). MR tier denotes if the protein-obesity trait association is supported in the sensitivity analyes and P_{het} tier denotes if the protein-obesity trait association have robust sexually heterogeneous effect estimates in the sensitivity analyses, as outlined in section 3.2.8.2. Effect estimates as change in sex-stratified SD units of the obesity trait per 1 unit higher sex-stratified SD of the protein. "-" denotes that the analysis was not performed, either because of the how the methods pipeline was created or because the protein-outcome combination did not surpass the P_{MR}<1% FDR threshold in that sex or P_{het}<1% FDR in the sex-comparison (Benjamini and Hochberg, 1995). Prx, number of proxies; M, men; W, women.

Appendix Table 3.6. Bidirectional MR results for the effect of obesity and cardiometabolic diseases on proteins with evidence of an association PMR<0.05.

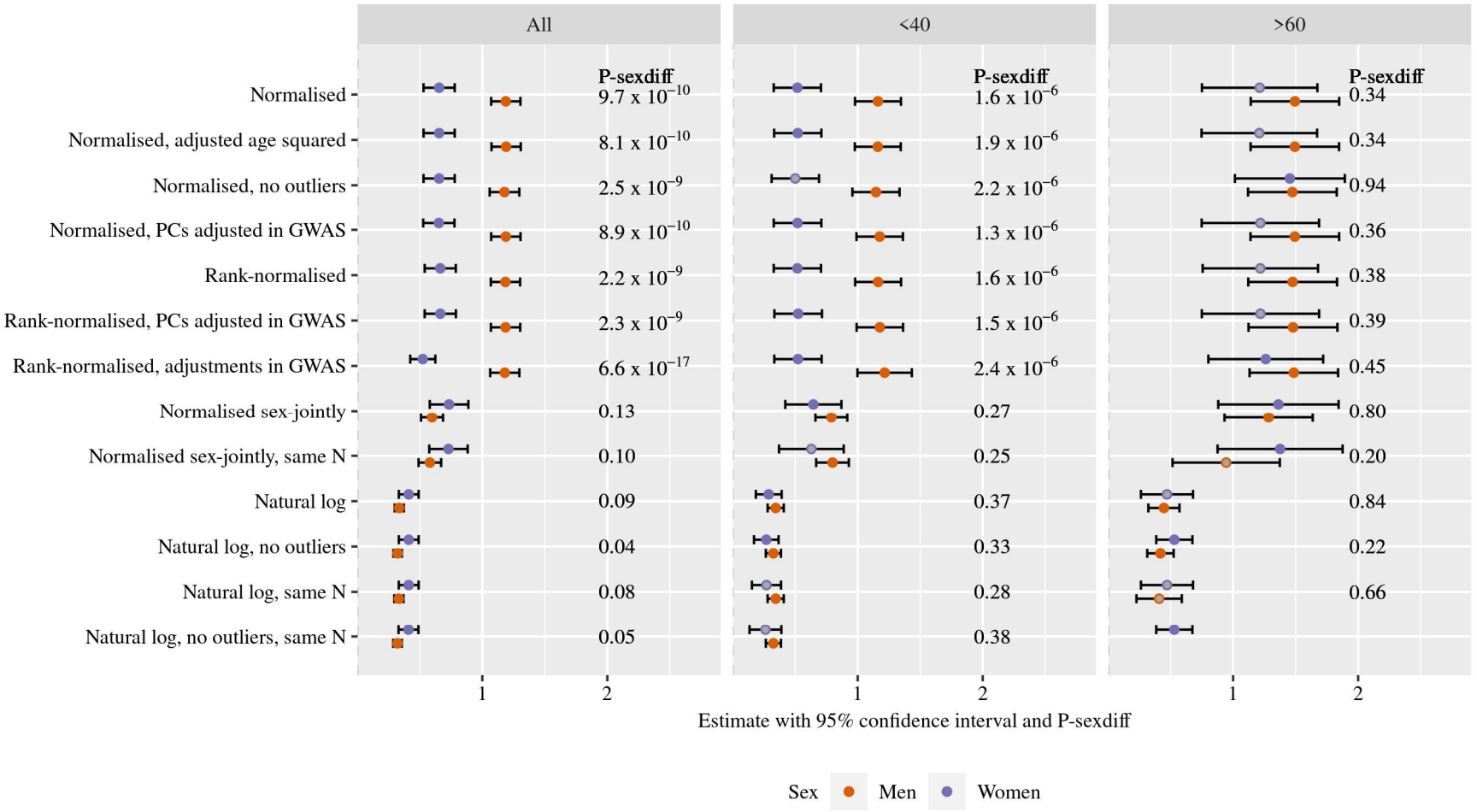
Exposure	Outcome, SOMAMER ID	Sex	Method	Instrumen N SNPs (proxies)	Estimate (95% CI)	P_{MR}
BMI	ACP1.3858.5.1	Women	MR Egger	493 (6)	0.711 (0.077,1.344)	0.028
BMI	ACP1.3858.5.1	Women	MR Egger intercept	493 (6)	-0.013 (-0.024,-0.001)	0.036
BMI	NELL1.6544.33.3	Women	Simple median	493 (6)	-0.315 (-0.606,-0.024)	0.034
BMI	NELL1.6544.33.3	Women	Weighted median	493 (6)	-0.350 (-0.667,-0.033)	0.031
BMI	NPTX1.9256.78.3	Women	Simple median	493 (6)	-0.307 (-0.593,-0.021)	0.035
BMI	NPTX1.9256.78.3	Women	Weighted median	493 (6)	-0.506 (-0.826,-0.187)	0.002
BMI	NPTX1.9256.78.3	Women	IVW	493 (6)	-0.274 (-0.464,-0.084)	0.005
BMI	NPTX1.9256.78.3	Women	MR Egger	493 (6)	-0.716 (-1.269,-0.163)	0.011
BMI	PAM.5620.13.3	Men	IVW	374 (4)	0.313 (0.120,0.506)	0.001
BMI	PAM.5620.13.3	Men	MR Egger	374 (4)	0.565 (0.015,1.114)	0.044
BMI	SCG3.7957.2.3	Women	Simple median	493 (6)	-0.458 (-0.747,-0.170)	0.002
BMI	SCG3.7957.2.3	Women	Weighted median	493 (6)	-0.369 (-0.682,-0.056)	0.021
BMI	SCG3.7957.2.3	Women	IVW	493 (6)	-0.447 (-0.640,-0.254)	5.8E-06
BMI	WFIKKN2.13408.23.3	Women	Simple median	493 (6)	-0.394 (-0.686,-0.103)	0.008
BMI	WFIKKN2.13408.23.3	Women	Weighted median	493 (6)	-0.323 (-0.629,-0.018)	0.038
BMI	WFIKKN2.13408.23.3	Women	IVW	493 (6)	-0.313 (-0.503,-0.122)	0.001
T2D, probable cases	PAM.5620.13.3	Men	MR Egger	29 (0)	-0.317 (-0.588,-0.046)	0.022
WHR	ATP1B2.7218.87.3	Women	Simple median	354 (2)	-0.367 (-0.665,-0.069)	0.016
WHR	ATP1B2.7218.87.3	Women	Weighted median	354 (2)	-0.394 (-0.730,-0.058)	0.022
WHR	ATP1B2.7218.87.3	Women	IVW	354 (2)	-0.211 (-0.415,-0.008)	0.042
WHR	CFH.4159.130.1	Women	Simple median	354 (2)	0.525 (0.218,0.832)	8.0E-04
WHR	CFH.4159.130.1	Women	Weighted median	354 (2)	0.507 (0.182,0.833)	0.002
WHR	CFH.4159.130.1	Women	IVW	354 (2)	0.529 (0.325,0.733)	3.6E-07
WHR	CFH.4159.130.1	Women	MR Egger	354 (2)	0.600 (0.108,1.092)	0.017

Exposure	Outcome, SOMAMER ID	Sex	Method	Instrumen N SNPs (proxies)	Estimate (95% CI)	P_{MR}
WHR	QPCTL.8866.53.3	Women	Weighted median	354 (2)	-0.356 (-0.689,-0.022)	0.036
WHR	QPCTL.8866.53.3	Women	IVW	354 (2)	-0.274 (-0.472,-0.076)	0.007
WHR	RSPO3.13094.75.3	Women	MR Egger intercept	354 (2)	-0.011 (-0.022,-0.000)	0.047
WHR	TIMP4.6462.12.3	Women	Weighted median	354 (2)	-0.576 (-0.917,-0.236)	9.2E-04
WHR	TIMP4.6462.12.3	Women	IVW	354 (2)	-0.434 (-0.647,-0.222)	6.1E-05
WHR	TIMP4.6462.12.3	Women	MR Egger	354 (2)	-0.865 (-1.376,-0.354)	9.0E-04
WHRadjBMI	ADAMTS5.3168.8.2	Women	IVW	434 (3)	0.165 (0.004,0.326)	0.045
WHRadjBMI	AKR1C1.12618.50.3	Men	Simple median	149 (0)	0.462 (0.022,0.902)	0.039
WHRadjBMI	ASPN.6451.64.3	Women	Simple median	434 (3)	0.275 (0.019,0.531)	0.035
WHRadjBMI	ATP1B2.7218.87.3	Women	Weighted median	434 (3)	-0.311 (-0.597,-0.025)	0.033
WHRadjBMI	ATP1B2.7218.87.3	Women	MR Egger	434 (3)	-0.427 (-0.809,-0.044)	0.029
WHRadjBMI	CFH.4159.130.1	Women	Simple median	434 (3)	0.340 (0.079,0.602)	0.011
WHRadjBMI	CFH.4159.130.1	Women	Weighted median	434 (3)	0.351 (0.064,0.639)	0.017
WHRadjBMI	CFH.4159.130.1	Women	IVW	434 (3)	0.297 (0.125,0.469)	7.2E-04
WHRadjBMI	NPPB.7655.11.3	Women	IVW	434 (3)	-0.234 (-0.415,-0.053)	0.011
WHRadjBMI	RSPO3.13094.75.3	Men	MR Egger	149 (0)	1.350 (0.250,2.449)	0.016
WHRadjBMI	RSPO3.13094.75.3	Men	MR Egger intercept	149 (0)	-0.030 (-0.053,-0.006)	0.013
WHRadjBMI	TIMP4.6462.12.3	Women	Weighted median	434 (3)	-0.483 (-0.769,-0.196)	9.6E-04
WHRadjBMI	TIMP4.6462.12.3	Women	IVW	434 (3)	-0.312 (-0.480,-0.143)	2.9E-04
WHRadjBMI	TIMP4.6462.12.3	Women	MR Egger	434 (3)	-0.716 (-1.084,-0.348)	1.3E-04
WHRadjBMI	TIMP4.6462.12.3	Women	MR Egger intercept	434 (3)	0.011 (0.002,0.020)	0.015

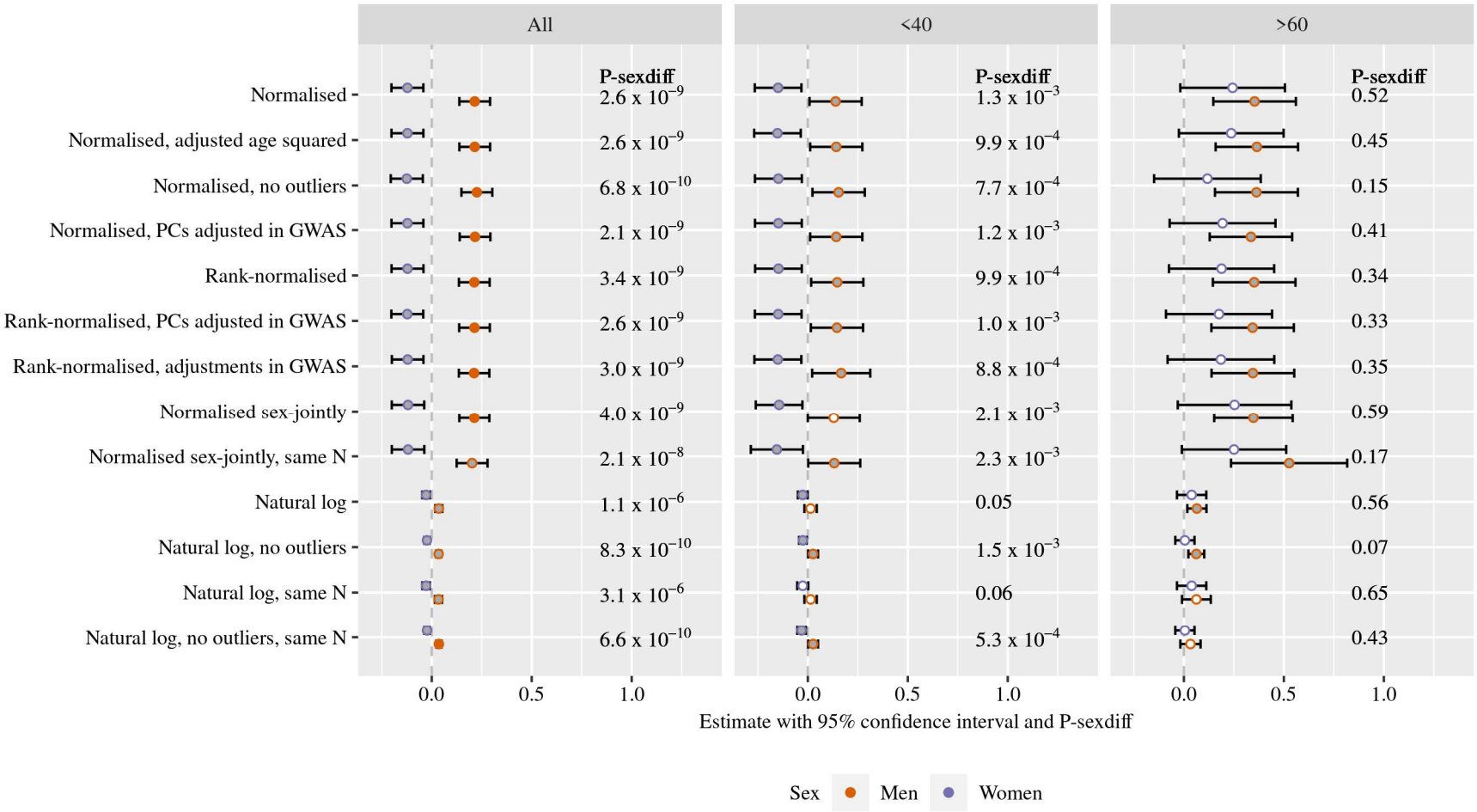
Only associations $P_{MR}<0.05$ are included in the table due to space restrictions. Only protein-outcome combinations where the protein had an effect on the outcome were investigated for a bidirectional relationship. Effect estimates in sex-specific SD units per 1 sex-specific SD for the obesity traits and per 1 sex-specific log(OR) for the cardiometabolic diseases.



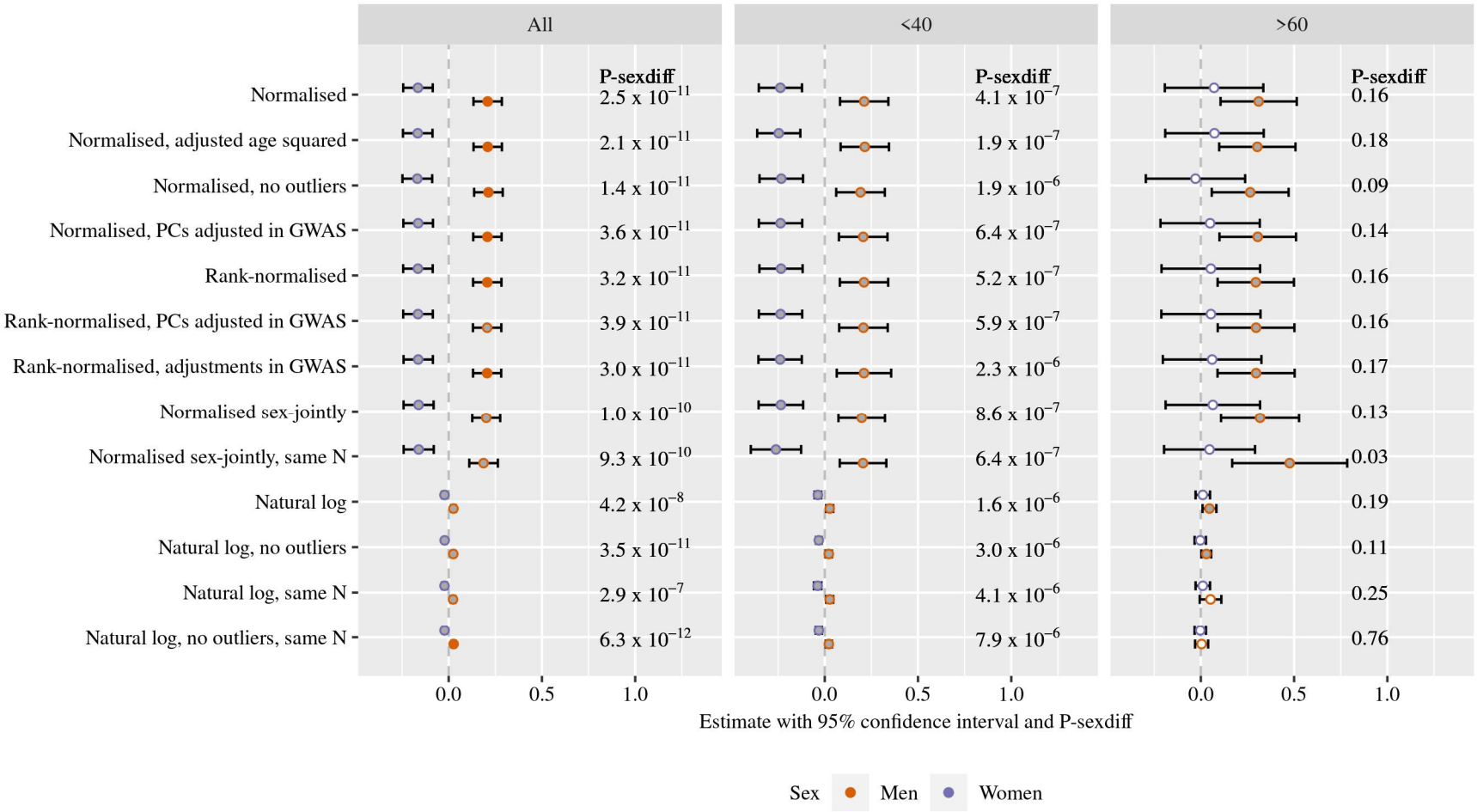
Appendix Figure 3.1. Observational correlations between aptamer pairs in a sex-combined analysis, using a different rank-transformation approach. For this analysis, all protein values were first RINTed, before a linear regression was performed and the residuals RINTed a second time (Bliss, 1967; Pain, Dudbridge and Ronald, 2018). Only aptamer pairs correlated $P_{correlation}$<1% FDR have their correlation r values colour-coded; non-significant aptamer-aptamer correlations are coded grey (Benjamini and Hochberg, 1995).



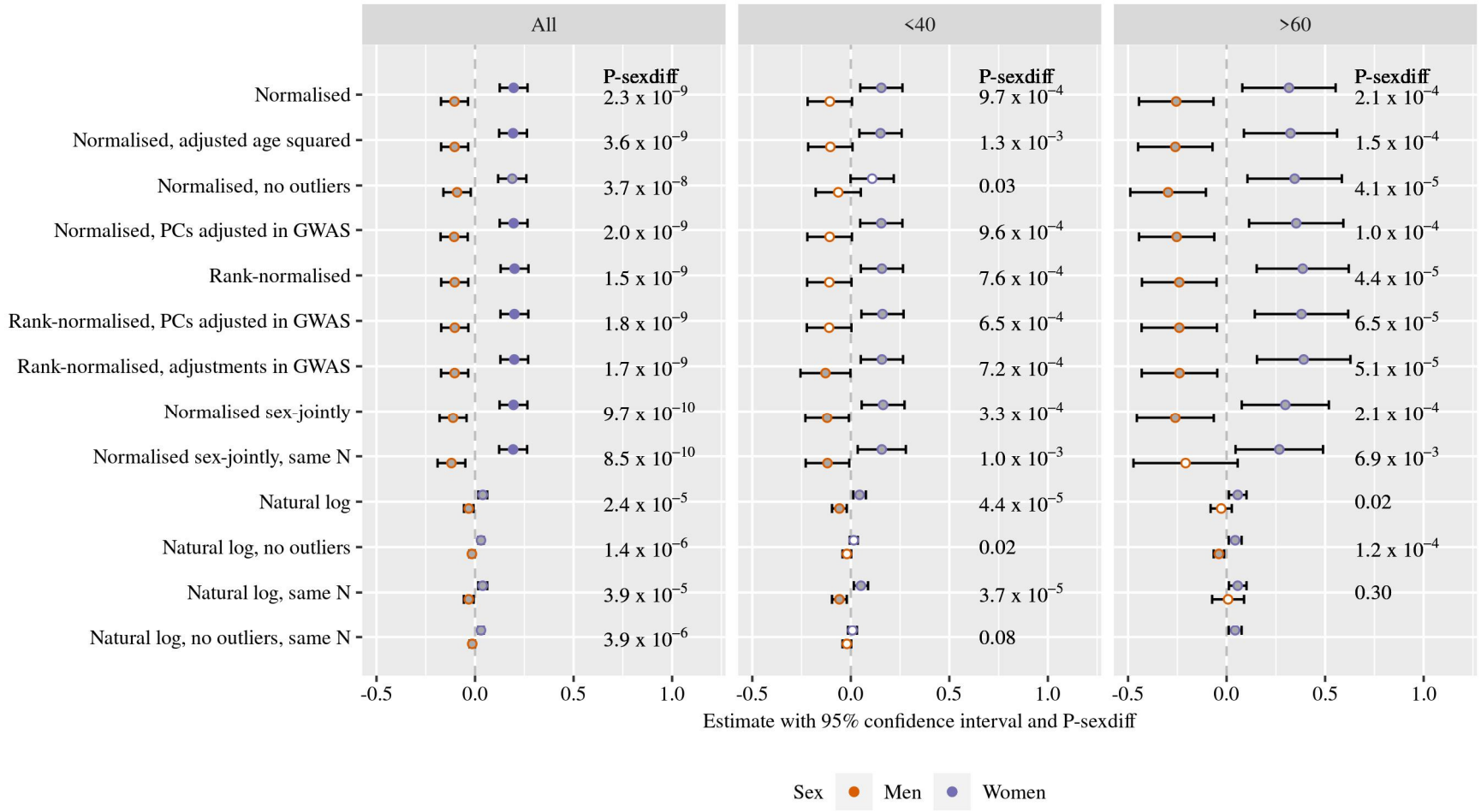
Appendix Figure 3.2. Estimated higher sex-stratified SDs of luteinizing hormone (measured with the CGA.LHB.2953.31.2 aptamer) per rs78537284 G-allele, using different protein transformation approaches, adjustment protocols, and in different age-strata, as outlined in section 3.2.9.4.1. Filled point estimates indicate P_{GWAS}<1% FDR in that analysis, grey point estimates indicate P_{GWAS}<0.05, white point estimates indicated P_{GWAS}≥0.05 (Benjamini and Hochberg, 1995).



Appendix Figure 3.3. Estimated higher sex-stratified SDs of angiopoietin-4 (measured with the ANGPT4.2500.2.3 aptamer) per rs35185467 C-allele using different protein transformation approaches, adjustment protocols, and in different age-strata, as outlined in section 3.2.9.4.1. Filled point estimates indicate P_{GWAS}<1% FDR in that analysis, grey point estimates indicate P_{GWAS}<0.05, white point estimates indicated P_{GWAS}≥0.05 (Benjamini and Hochberg, 1995).



Appendix Figure 3.4. Estimated higher sex-stratified SDs of ephrin-B3 (measured with the EFNB3.2514.65.3 aptamer) per rs17065165 A-allele using different protein transformation approaches, adjustment protocols, and in different age-strata, as outlined in section 3.2.9.4.1. Filled point estimates indicate P_{GWAS}<1% FDR in that analysis, grey point estimates indicate P_{GWAS}<0.05, white point estimates indicated P_{GWAS}≥0.05 (Benjamini and Hochberg, 1995).



Appendix Figure 3.5. Estimated higher sex-stratified SDs of zinc finger protein SNAI2 (measured with the SNAI2.10014.31.3 aptamer) per rs3920702 G-allele using different protein transformation approaches, adjustment protocols, and in different age-strata, as outlined in section 3.2.9.4.1. Filled point estimates indicate P_{GWAS}<1% FDR in that analysis, grey point estimates indicate P_{GWAS}<0.05, white point estimates indicated P_{GWAS}≥0.05 (Benjamini and Hochberg, 1995).

Appendix 4. Supplement to Chapter 4

Appendix Table 4.1. MR estimates of the effect of the obesity traits on miscarriage and stillbirth.

Exposure	N SNPs analysis (proxies)	Method	OR (95% CI)	P
BMI	39 (0)	IVW	0.86 (0.68-1.09)	0.21
BMI	39 (0)	Simple median	0.86 (0.62-1.20)	0.38
BMI	39 (0)	Weighted median	0.78 (0.56-1.10)	0.16
BMI	39 (0)	MR Egger	0.85 (0.44-1.63)	0.62
WHR	22 (0)	IVW	0.77 (0.59-1.01)	0.06
WHR	22 (0)	Simple median	0.80 (0.55-1.16)	0.24
WHR	22 (0)	Weighted median	0.81 (0.56-1.17)	0.25
WHR	22 (0)	MR Egger	0.93 (0.22-3.97)	0.93
WHRadjBMI	33 (0)	IVW	0.86 (0.70-1.06)	0.17
WHRadjBMI	33 (0)	Simple median	0.89 (0.66-1.20)	0.44
WHRadjBMI	33 (0)	Weighted median	0.80 (0.60-1.08)	0.15
WHRadjBMI	33 (0)	MR Egger	0.59 (0.27-1.32)	0.20

Effect estimates per 1 women-specific SD unit higher obesity trait.

Appendix Table 4.2. MR estimates of the effect of PCOS on obesity.

Outcome	N SNPs analysis (proxies)	Method	Estimate (95% CI)	P
BMI	13 (0)	IVW	-0.005 (-0.033, 0.022)	0.71
BMI	13 (0)	Simple median	0.003 (-0.018, 0.023)	0.81
BMI	13 (0)	Weighted median	0.001 (-0.019, 0.020)	0.93
BMI	13 (0)	MR Egger	0.032 (-0.096, 0.161)	0.62
BMI	13 (0)	MR Egger intercept	-0.005 (-0.021, 0.011)	0.55
WHR	13 (0)	IVW	0.014 (-0.014, 0.042)	0.33
WHR	13 (0)	Simple median	0.024 (-0.000, 0.047)	0.05
WHR	13 (0)	Weighted median	0.014 (-0.008, 0.036)	0.21
WHR	13 (0)	MR Egger	0.039 (-0.092, 0.171)	0.56
WHR	13 (0)	MR Egger intercept	-0.003 (-0.020, 0.013)	0.70
WHRadjBMI	13 (0)	IVW	0.020 (-0.004, 0.044)	0.10
WHRadjBMI	13 (0)	Simple median	0.014 (-0.008, 0.036)	0.22
WHRadjBMI	13 (0)	Weighted median	0.013 (-0.007, 0.034)	0.20
WHRadjBMI	13 (0)	MR Egger	0.042 (-0.070, 0.155)	0.46
WHRadjBMI	13 (0)	MR Egger intercept	-0.003 (-0.017, 0.011)	0.69

Effect estimates in women-specific SDs of each obesity trait per 1 log(OR) of PCOS.

Appendix Table 4.3. Fixed effect IVW meta-analysis for the obesity traits using the SNPs associated with miscarriage.

Outcome	N SNPs analysis (proxies)	Estimate (95% CI)	P
BMI	2 (1)	0.006 (-0.009, 0.020)	1.00
WHR	2 (1)	0.001 (-0.015, 0.016)	1.00
WHRadjBMI	2 (1)	-0.003 (-0.020, 0.013)	1.00

Estimated higher women-specific SDs per additional risk allele for recurrent miscarriage.

Appendix Table 4.4. Coloc PPs for colocalisation for all gene/protein and tissue combinations with evidence of colocalisation (PP>0.50) with PCOS in any Coloc analysis.

PCOS SNP	Gene/Protein	QTL dataset	Tissue	Coloc, 2 Mb, Main	Coloc, 2 Mb, Without-23	Coloc, 400 kb, Main	Coloc, 400 kb, Without-23	Coloc, independent region, Main	Coloc, independent region, Without-23
rs10739076	*RCL1*	eQTLgen	Blood expression levels	0.00	0.00	0.82	0.00	-	-
rs11031005	*ARL14EP*	GTEx	Testis	0.42	0.80	0.42	0.80	-	-
rs11031005	FSH	INTERVAL	Blood protein levels	0.76	0.67	0.76	0.68	0.76	0.68
rs11225154	*TMEM123*	GTEx	Brain Cortex	0.45	0.00	0.54	0.00	-	-
rs13164856	*RAD50*	GTEx	Adipose Visceral Omentum	0.46	0.11	0.47	0.13	-	-
rs13164856	*RAD50*	GTEx	Heart Left Ventricle	0.53	0.15	0.54	0.17	-	-
rs2271194	*ERBB3*	GTEx	Brain Spinal cord cervical c-1	0.51	0.11	0.55	0.14	-	-
rs2271194	*ERBB3*	eQTLgen	Blood expression levels	0.94	0.44	0.94	0.48	-	-
rs2271194	*GDF11*	eQTLgen	Blood expression levels	0.53	0.49	0.53	0.54	-	-
rs2271194	*IKZF4*	GTEx	Adipose Subcutaneous	0.60	0.20	0.61	0.23	-	-
rs2271194	*IKZF4*	GTEx	Esophagus Mucosa	0.90	0.42	0.90	0.45	-	-
rs2271194	*IKZF4*	GTEx	Pancreas	0.44	0.09	0.51	0.12	-	-
rs2271194	*IKZF4*	GTEx	Skin Not Sun Exposed Suprapubic	0.50	0.12	0.54	0.16	-	-
rs2271194	*RPS26*	GTEx	Adipose Subcutaneous	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Adipose Visceral Omentum	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Adrenal Gland	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Artery Aorta	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Artery Coronary	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Artery Tibial	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Brain Amygdala	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Brain Anterior cingulate cortex BA24	0.80	0.56	0.80	0.60	-	-

PCOS SNP	Gene/Protein	QTL dataset	Tissue	Coloc, 2 Mb, Main	Coloc, 2 Mb, Without-23	Coloc, 400 kb, Main	Coloc, 400 kb, Without-23	Coloc, independent region, Main	Coloc, independent region, Without-23
rs2271194	*RPS26*	GTEx	Brain Caudate basal ganglia	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Brain Cerebellar Hemisphere	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Brain Cerebellum	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Brain Cortex	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Brain Frontal Cortex BA9	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Brain Hippocampus	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Brain Hypothalamus	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Brain Nucleus accumbens basal ganglia	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Brain Putamen basal ganglia	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Brain Spinal cord cervical c-1	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Brain Substantia nigra	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Breast Mammary Tissue	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Cells EBV-transformed lymphocytes	0.70	0.45	0.70	0.49	-	-
rs2271194	*RPS26*	GTEx	Cells Transformed fibroblasts	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Colon Sigmoid	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Colon Transverse	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Esophagus Gastroesophageal Junction	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Esophagus Mucosa	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Esophagus Muscularis	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Heart Atrial Appendage	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Heart Left Ventricle	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Liver	0.80	0.56	0.80	0.60	-	-

PCOS SNP	Gene/Protein	QTL dataset	Tissue	Coloc, 2 Mb, Main	Coloc, 2 Mb, Without-23	Coloc, 400 kb, Main	Coloc, 400 kb, Without-23	Coloc, independent region, Main	Coloc, independent region, Without-23
rs2271194	*RPS26*	GTEx	Lung	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Minor Salivary Gland	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Muscle Skeletal	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Nerve Tibial	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Ovary	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Pancreas	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Pituitary	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Prostate	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Skin Not Sun Exposed Suprapubic	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Skin Sun Exposed Lower leg	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Small Intestine Terminal Ileum	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Spleen	0.80	0.57	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Stomach	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Testis	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Thyroid	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Uterus	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Vagina	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	GTEx	Whole Blood	0.80	0.56	0.80	0.60	-	-
rs2271194	*RPS26*	eQTLgen	Blood expression levels	0.91	0.46	0.91	0.50	-	-
rs2271194	*SUOX*	GTEx	Adipose Subcutaneous	0.78	0.31	0.78	0.34	-	-
rs2271194	*SUOX*	GTEx	Adrenal Gland	0.82	0.29	0.84	0.34	-	-
rs2271194	*SUOX*	GTEx	Artery Aorta	0.85	0.54	0.85	0.57	-	-
rs2271194	*SUOX*	GTEx	Artery Tibial	0.51	0.07	0.51	0.08	-	-

PCOS SNP	Gene/Protein	QTL dataset	Tissue	Coloc, 2 Mb, Main	Coloc, 2 Mb, Without-23	Coloc, 400 kb, Main	Coloc, 400 kb, Without-23	Coloc, independent region, Main	Coloc, independent region, Without-23
rs2271194	*SUOX*	GTEx	Brain Amygdala	0.83	0.42	0.84	0.46	-	-
rs2271194	*SUOX*	GTEx	Brain Anterior cingulate cortex BA24	0.84	0.41	0.85	0.45	-	-
rs2271194	*SUOX*	GTEx	Brain Caudate basal ganglia	0.86	0.47	0.86	0.50	-	-
rs2271194	*SUOX*	GTEx	Brain Cerebellar Hemisphere	0.74	0.47	0.74	0.50	-	-
rs2271194	*SUOX*	GTEx	Brain Cerebellum	0.62	0.29	0.62	0.32	-	-
rs2271194	*SUOX*	GTEx	Brain Cortex	0.75	0.24	0.75	0.27	-	-
rs2271194	*SUOX*	GTEx	Brain Frontal Cortex BA9	0.64	0.12	0.68	0.16	-	-
rs2271194	*SUOX*	GTEx	Brain Hippocampus	0.79	0.48	0.79	0.51	-	-
rs2271194	*SUOX*	GTEx	Brain Hypothalamus	0.90	0.49	0.90	0.53	-	-
rs2271194	*SUOX*	GTEx	Brain Nucleus accumbens basal ganglia	0.77	0.28	0.78	0.31	-	-
rs2271194	*SUOX*	GTEx	Brain Putamen basal ganglia	0.63	0.11	0.63	0.13	-	-
rs2271194	*SUOX*	GTEx	Breast Mammary Tissue	0.75	0.30	0.75	0.33	-	-
rs2271194	*SUOX*	GTEx	Cells Transformed fibroblasts	0.51	0.30	0.52	0.34	-	-
rs2271194	*SUOX*	GTEx	Colon Sigmoid	0.82	0.53	0.82	0.57	-	-
rs2271194	*SUOX*	GTEx	Colon Transverse	0.89	0.32	0.89	0.35	-	-
rs2271194	*SUOX*	GTEx	Esophagus Gastroesophageal Junction	0.76	0.23	0.76	0.26	-	-
rs2271194	*SUOX*	GTEx	Esophagus Mucosa	0.52	0.07	0.52	0.08	-	-
rs2271194	*SUOX*	GTEx	Esophagus Muscularis	0.51	0.07	0.52	0.09	-	-
rs2271194	*SUOX*	GTEx	Heart Atrial Appendage	0.71	0.34	0.71	0.37	-	-
rs2271194	*SUOX*	GTEx	Heart Left Ventricle	0.50	0.06	0.50	0.07	-	-
rs2271194	*SUOX*	GTEx	Lung	0.63	0.09	0.63	0.10	-	-
rs2271194	*SUOX*	GTEx	Muscle Skeletal	0.86	0.32	0.86	0.35	-	-

PCOS SNP	Gene/Protein	QTL dataset	Tissue	Coloc, 2 Mb, Main	Coloc, 2 Mb, Without-23	Coloc, 400 kb, Main	Coloc, 400 kb, Without-23	Coloc, independent region, Main	Coloc, independent region, Without-23
rs2271194	*SUOX*	GTEx	Nerve Tibial	0.64	0.09	0.64	0.10	-	-
rs2271194	*SUOX*	GTEx	Ovary	0.57	0.09	0.57	0.10	-	-
rs2271194	*SUOX*	GTEx	Pancreas	0.61	0.08	0.61	0.10	-	-
rs2271194	*SUOX*	GTEx	Pituitary	0.92	0.45	0.92	0.49	-	-
rs2271194	*SUOX*	GTEx	Prostate	0.51	0.07	0.51	0.08	-	-
rs2271194	*SUOX*	GTEx	Skin Not Sun Exposed Suprapubic	0.76	0.50	0.76	0.54	-	-
rs2271194	*SUOX*	GTEx	Small Intestine Terminal Ileum	0.70	0.13	0.70	0.14	-	-
rs2271194	*SUOX*	GTEx	Spleen	0.53	0.06	0.53	0.06	-	-
rs2271194	*SUOX*	GTEx	Stomach	0.54	0.08	0.54	0.10	-	-
rs2271194	*SUOX*	GTEx	Testis	0.76	0.48	0.76	0.52	-	-
rs2271194	*SUOX*	GTEx	Thyroid	0.62	0.08	0.62	0.10	-	-
rs2271194	*SUOX*	GTEx	Whole Blood	0.61	0.14	0.62	0.16	-	-
rs2271194	*SUOX*	eQTLgen	Blood expression levels	0.70	0.29	0.71	0.33	-	-
rs2271194	*ZC3H10*	GTEx	Skin Sun Exposed Lower leg	0.49	0.08	0.53	0.10	-	-
rs7563201	*ZFP36L2*	eQTLgen	Blood expression levels	0.93	0.01	0.93	0.01	-	-
rs7864171	*C9orf3*	GTEx	Heart Atrial Appendage	0.60	0.13	0.82	0.30	-	-
rs804279	*C8orf49*	GTEx	Stomach	0.87	0.19	0.88	0.20	-	-
rs804279	*NEIL2*	GTEx	Adrenal Gland	0.60	0.42	0.61	0.43	-	-
rs804279	*NEIL2*	GTEx	Cells EBV-transformed lymphocytes	0.73	0.75	0.73	0.75	-	-
rs804279	*NEIL2*	GTEx	Cells Transformed fibroblasts	0.60	0.37	0.60	0.37	-	-

Analyses performed using both the main PCOS dataset (with summary statistics from the maximum sample size for each SNP) and the Without-23 PCOS dataset (with summary statistics from approximately similar sample sizes for each SNP) as well as using different region sizes (2 Mb, 400 kb, and, if full GWAS data available, also "independent regions" as per Berisa and Pickrell (2016)). "-" signifies that the analyses could not be performed due to missing

data to perform the analysis. Table adapted (changed text, layout and included information) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

Appendix Table 4.5. Power results (the sum of the PPs for no colocalisation (hypothesis 3) and colocalisation (hypothesis 4) for all gene/protein and tissue combinations with evidence of colocalisation (PP>0.50) with PCOS in any Coloc analysis.

PCOS SNP	Gene/Protein	QTL dataset	Tissue	Coloc, 2 Mb, Main	Coloc, 2 Mb, Without-23	Coloc, 400 kb, Main	Coloc, 400 kb, Without-23	Coloc, independent regions, Main	Coloc, independent regions, Without-23
rs10739076	*RCL1*	eQTLgen	Blood expression levels	0.98	0.32	0.83	0.01	-	-
rs11031005	*ARL14EP*	GTEx	Testis	1.00	1.00	1.00	1.00	-	-
rs11031005	FSH	INTERVAL	Blood protein levels	0.99	0.98	0.99	0.98	0.99	0.98
rs11225154	*TMEM123*	GTEx	Brain Cortex	0.65	0.10	0.57	0.01	-	-
rs13164856	*RAD50*	GTEx	Adipose Visceral Omentum	0.88	0.30	0.87	0.23	-	-
rs13164856	*RAD50*	GTEx	Heart Left Ventricle	0.98	0.37	0.98	0.29	-	-
rs2271194	*ERBB3*	GTEx	Brain Spinal cord cervical c-1	0.72	0.28	0.70	0.25	-	-
rs2271194	*ERBB3*	eQTLgen	Blood expression levels	1.00	0.70	1.00	0.68	-	-
rs2271194	*GDF11*	eQTLgen	Blood expression levels	1.00	0.74	1.00	0.71	-	-
rs2271194	*IKZF4*	GTEx	Adipose Subcutaneous	0.94	0.51	0.93	0.47	-	-
rs2271194	*IKZF4*	GTEx	Esophagus Mucosa	1.00	0.67	1.00	0.65	-	-
rs2271194	*IKZF4*	GTEx	Pancreas	0.73	0.30	0.69	0.24	-	-
rs2271194	*IKZF4*	GTEx	Skin Not Sun Exposed Suprapubic	0.74	0.31	0.72	0.28	-	-
rs2271194	*RPS26*	GTEx	Adipose Subcutaneous	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Adipose Visceral Omentum	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Adrenal Gland	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Artery Aorta	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Artery Coronary	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Artery Tibial	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Brain Amygdala	1.00	0.76	1.00	0.74	-	-

PCOS SNP	Gene/Protein	QTL dataset	Tissue	Coloc, 2 Mb, Main	Coloc, 2 Mb, Without-23	Coloc, 400 kb, Main	Coloc, 400 kb, Without-23	Coloc, independent regions, Main	Coloc, independent regions, Without-23
rs2271194	*RPS26*	GTEx	Brain Anterior cingulate cortex BA24	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Brain Caudate basal ganglia	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Brain Cerebellar Hemisphere	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Brain Cerebellum	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Brain Cortex	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Brain Frontal Cortex BA9	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Brain Hippocampus	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Brain Hypothalamus	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Brain Nucleus accumbens basal ganglia	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Brain Putamen basal ganglia	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Brain Spinal cord cervical c-1	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Brain Substantia nigra	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Breast Mammary Tissue	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Cells EBV-transformed lymphocytes	1.00	0.70	1.00	0.67	-	-
rs2271194	*RPS26*	GTEx	Cells Transformed fibroblasts	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Colon Sigmoid	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Colon Transverse	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Esophagus Gastroesophageal Junction	1.00	0.76	1.00	0.74	-	-

PCOS SNP	Gene/Protein	QTL dataset	Tissue	Coloc, 2 Mb, Main	Coloc, 2 Mb, Without-23	Coloc, 400 kb, Main	Coloc, 400 kb, Without-23	Coloc, independent regions, Main	Coloc, independent regions, Without-23
rs2271194	*RPS26*	GTEx	Esophagus Mucosa	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Esophagus Muscularis	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Heart Atrial Appendage	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Heart Left Ventricle	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Liver	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Lung	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Minor Salivary Gland	1.00	0.75	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Muscle Skeletal	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Nerve Tibial	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Ovary	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Pancreas	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Pituitary	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Prostate	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Skin Not Sun Exposed Suprapubic	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Skin Sun Exposed Lower leg	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Small Intestine Terminal Ileum	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Spleen	1.00	0.75	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Stomach	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Testis	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Thyroid	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Uterus	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	GTEx	Vagina	1.00	0.76	1.00	0.74	-	-

PCOS SNP	Gene/Protein	QTL dataset	Tissue	Coloc, 2 Mb, Main	Coloc, 2 Mb, Without-23	Coloc, 400 kb, Main	Coloc, 400 kb, Without-23	Coloc, independent regions, Main	Coloc, independent regions, Without-23
rs2271194	*RPS26*	GTEx	Whole Blood	1.00	0.76	1.00	0.74	-	-
rs2271194	*RPS26*	eQTLgen	Blood expression levels	1.00	0.71	1.00	0.69	-	-
rs2271194	*SUOX*	GTEx	Adipose Subcutaneous	1.00	0.62	1.00	0.58	-	-
rs2271194	*SUOX*	GTEx	Adrenal Gland	0.95	0.52	0.95	0.51	-	-
rs2271194	*SUOX*	GTEx	Artery Aorta	1.00	0.74	1.00	0.73	-	-
rs2271194	*SUOX*	GTEx	Artery Tibial	1.00	0.49	1.00	0.42	-	-
rs2271194	*SUOX*	GTEx	Brain Amygdala	0.99	0.67	0.99	0.65	-	-
rs2271194	*SUOX*	GTEx	Brain Anterior cingulate cortex BA24	1.00	0.67	1.00	0.64	-	-
rs2271194	*SUOX*	GTEx	Brain Caudate basal ganglia	1.00	0.71	1.00	0.68	-	-
rs2271194	*SUOX*	GTEx	Brain Cerebellar Hemisphere	1.00	0.71	1.00	0.68	-	-
rs2271194	*SUOX*	GTEx	Brain Cerebellum	1.00	0.61	1.00	0.57	-	-
rs2271194	*SUOX*	GTEx	Brain Cortex	1.00	0.58	1.00	0.54	-	-
rs2271194	*SUOX*	GTEx	Brain Frontal Cortex BA9	0.79	0.29	0.78	0.26	-	-
rs2271194	*SUOX*	GTEx	Brain Hippocampus	1.00	0.71	1.00	0.69	-	-
rs2271194	*SUOX*	GTEx	Brain Hypothalamus	1.00	0.72	1.00	0.70	-	-
rs2271194	*SUOX*	GTEx	Brain Nucleus accumbens basal ganglia	0.99	0.59	0.99	0.55	-	-
rs2271194	*SUOX*	GTEx	Brain Putamen basal ganglia	1.00	0.51	1.00	0.45	-	-
rs2271194	*SUOX*	GTEx	Breast Mammary Tissue	1.00	0.61	1.00	0.57	-	-
rs2271194	*SUOX*	GTEx	Cells Transformed fibroblasts	0.98	0.60	0.98	0.57	-	-
rs2271194	*SUOX*	GTEx	Colon Sigmoid	1.00	0.74	1.00	0.72	-	-

PCOS SNP	Gene/Protein	QTL dataset	Tissue	Coloc, 2 Mb, Main	Coloc, 2 Mb, Without-23	Coloc, 400 kb, Main	Coloc, 400 kb, Without-23	Coloc, independent regions, Main	Coloc, independent regions, Without-23
rs2271194	*SUOX*	GTEx	Colon Transverse	1.00	0.63	1.00	0.59	-	-
rs2271194	*SUOX*	GTEx	Esophagus Gastroesophageal Junction	1.00	0.58	1.00	0.53	-	-
rs2271194	*SUOX*	GTEx	Esophagus Mucosa	1.00	0.49	1.00	0.41	-	-
rs2271194	*SUOX*	GTEx	Esophagus Muscularis	1.00	0.49	1.00	0.42	-	-
rs2271194	*SUOX*	GTEx	Heart Atrial Appendage	1.00	0.64	1.00	0.60	-	-
rs2271194	*SUOX*	GTEx	Heart Left Ventricle	1.00	0.49	1.00	0.41	-	-
rs2271194	*SUOX*	GTEx	Lung	1.00	0.50	1.00	0.43	-	-
rs2271194	*SUOX*	GTEx	Muscle Skeletal	1.00	0.62	1.00	0.58	-	-
rs2271194	*SUOX*	GTEx	Nerve Tibial	1.00	0.50	1.00	0.43	-	-
rs2271194	*SUOX*	GTEx	Ovary	0.99	0.49	0.99	0.42	-	-
rs2271194	*SUOX*	GTEx	Pancreas	1.00	0.50	1.00	0.43	-	-
rs2271194	*SUOX*	GTEx	Pituitary	1.00	0.70	1.00	0.67	-	-
rs2271194	*SUOX*	GTEx	Prostate	1.00	0.49	1.00	0.41	-	-
rs2271194	*SUOX*	GTEx	Skin Not Sun Exposed Suprapubic	1.00	0.73	1.00	0.71	-	-
rs2271194	*SUOX*	GTEx	Small Intestine Terminal Ileum	1.00	0.52	1.00	0.46	-	-
rs2271194	*SUOX*	GTEx	Spleen	1.00	0.48	1.00	0.41	-	-
rs2271194	*SUOX*	GTEx	Stomach	1.00	0.50	1.00	0.43	-	-
rs2271194	*SUOX*	GTEx	Testis	1.00	0.71	1.00	0.69	-	-
rs2271194	*SUOX*	GTEx	Thyroid	1.00	0.50	1.00	0.43	-	-
rs2271194	*SUOX*	GTEx	Whole Blood	1.00	0.53	1.00	0.47	-	-
rs2271194	*SUOX*	eQTLgen	Blood expression levels	1.00	0.63	1.00	0.58	-	-

PCOS SNP	Gene/Protein	QTL dataset	Tissue	Coloc, 2 Mb, Main	Coloc, 2 Mb, Without-23	Coloc, 400 kb, Main	Coloc, 400 kb, Without-23	Coloc, independent regions, Main	Coloc, independent regions, Without-23
rs2271194	*ZC3H10*	GTEx	Skin Sun Exposed Lower leg	0.68	0.23	0.65	0.19	-	-
rs7563201	*ZFP36L2*	eQTLgen	Blood expression levels	1.00	0.28	1.00	0.08	-	-
rs7864171	*C9orf3*	GTEx	Heart Atrial Appendage	1.00	0.43	1.00	0.41	-	-
rs804279	*C8orf49*	GTEx	Stomach	1.00	0.99	1.00	0.99	-	-
rs804279	*NEIL2*	GTEx	Adrenal Gland	0.95	0.92	0.95	0.91	-	-
rs804279	*NEIL2*	GTEx	Cells EBV-transformed lymphocytes	0.98	0.98	0.98	0.98	-	-
rs804279	*NEIL2*	GTEx	Cells Transformed fibroblasts	1.00	0.99	1.00	0.99	-	-

Analyses performed using both the main PCOS dataset (with summary statistics from the maximum sample size for each SNP) and the Without-23 PCOS dataset (with summary statistics from approximately similar sample sizes for each SNP) as well as using different region sizes (2 Mb, 400 kb, and, if full GWAS data available, also "independent regions" as per Berisa and Pickrell (2016)). "-" signifies that the analyses could not be performed due to missing data to perform the analysis. Table adapted (changed text, layout and included information) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

Appendix Table 4.6. PPs computed using HyPrColoc for all gene/protein and tissue combinations with evidence of colocalisation (PP>0.50) with PCOS in any Coloc analysis.

PCOS SNP	Gene/Protein	QTL dataset	Tissue	HyPrColoc, 2 Mb, Main	HyPrColoc, 2 Mb, Without-23	HyPrColoc, independent regions, Main	HyPrColoc, independent regions, Without-23
rs10739076	*RCL1*	eQTLgen	Blood expression levels	NS	NS	-	-
rs11031005	*ARL14EP*	GTEx	Testis	NS	0.94	-	-
rs11031005	FSH	INTERVAL	Blood protein levels	0.92	0.89	0.92	0.89
rs11225154	*TMEM123*	GTEx	Brain Cortex	NS	NS	-	-
rs13164856	*RAD50*	GTEx	Adipose Visceral Omentum	0.72	NS	-	-
rs13164856	*RAD50*	GTEx	Heart Left Ventricle	0.81	NS	-	-
rs2271194	*ERBB3*	GTEx	Brain Spinal cord cervical c-1	NS	NS	-	-
rs2271194	*ERBB3*	eQTLgen	Blood expression levels	0.98	NS	-	-
rs2271194	*GDF11*	eQTLgen	Blood expression levels	0.82	NS	-	-
rs2271194	*IKZF4*	GTEx	Adipose Subcutaneous	0.83	NS	-	-
rs2271194	*IKZF4*	GTEx	Esophagus Mucosa	0.97	NS	-	-
rs2271194	*IKZF4*	GTEx	Pancreas	NS	NS	-	-
rs2271194	*IKZF4*	GTEx	Skin Not Sun Exposed Suprapubic	NS	NS	-	-
rs2271194	*RPS26*	GTEx	Adipose Subcutaneous	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Adipose Visceral Omentum	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Adrenal Gland	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Artery Aorta	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Artery Coronary	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Artery Tibial	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Brain Amygdala	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Brain Anterior cingulate cortex BA24	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Brain Caudate basal ganglia	0.94	0.76	-	-

PCOS SNP	Gene/Protein	QTL dataset	Tissue	HyPrColoc, 2 Mb, Main	HyPrColoc, 2 Mb, Without-23	HyPrColoc, independent regions, Main	HyPrColoc, independent regions, Without-23
rs2271194	*RPS26*	GTEx	Brain Cerebellar Hemisphere	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Brain Cerebellum	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Brain Cortex	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Brain Frontal Cortex BA9	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Brain Hippocampus	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Brain Hypothalamus	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Brain Nucleus accumbens basal ganglia	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Brain Putamen basal ganglia	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Brain Spinal cord cervical c-1	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Brain Substantia nigra	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Breast Mammary Tissue	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Cells EBV-transformed lymphocytes	0.90	NS	-	-
rs2271194	*RPS26*	GTEx	Cells Transformed fibroblasts	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Colon Sigmoid	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Colon Transverse	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Esophagus Gastroesophageal Junction	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Esophagus Mucosa	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Esophagus Muscularis	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Heart Atrial Appendage	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Heart Left Ventricle	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Liver	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Lung	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Minor Salivary Gland	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Muscle Skeletal	0.94	0.76	-	-

PCOS SNP	Gene/Protein	QTL dataset	Tissue	HyPrColoc, 2 Mb, Main	HyPrColoc, 2 Mb, Without-23	HyPrColoc, independent regions, Main	HyPrColoc, independent regions, Without-23
rs2271194	*RPS26*	GTEx	Nerve Tibial	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Ovary	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Pancreas	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Pituitary	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Prostate	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Skin Not Sun Exposed Suprapubic	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Skin Sun Exposed Lower leg	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Small Intestine Terminal Ileum	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Spleen	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Stomach	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Testis	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Thyroid	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Uterus	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Vagina	0.94	0.76	-	-
rs2271194	*RPS26*	GTEx	Whole Blood	0.94	0.76	-	-
rs2271194	*RPS26*	eQTLgen	Blood expression levels	0.93	0.75	-	-
rs2271194	*SUOX*	GTEx	Adipose Subcutaneous	0.93	NS	-	-
rs2271194	*SUOX*	GTEx	Adrenal Gland	0.93	NS	-	-
rs2271194	*SUOX*	GTEx	Artery Aorta	0.96	0.74	-	-
rs2271194	*SUOX*	GTEx	Artery Tibial	0.80	NS	-	-
rs2271194	*SUOX*	GTEx	Brain Amygdala	0.95	NS	-	-
rs2271194	*SUOX*	GTEx	Brain Anterior cingulate cortex BA24	0.95	NS	-	-
rs2271194	*SUOX*	GTEx	Brain Caudate basal ganglia	0.96	NS	-	-
rs2271194	*SUOX*	GTEx	Brain Cerebellar Hemisphere	0.92	NS	-	-

PCOS SNP	Gene/Protein	QTL dataset	Tissue	HyPrColoc, 2 Mb, Main	HyPrColoc, 2 Mb, Without-23	HyPrColoc, independent regions, Main	HyPrColoc, independent regions, Without-23
rs2271194	*SUOX*	GTEx	Brain Cerebellum	0.87	NS	-	-
rs2271194	*SUOX*	GTEx	Brain Cortex	0.92	NS	-	-
rs2271194	*SUOX*	GTEx	Brain Frontal Cortex BA9	0.81	NS	-	-
rs2271194	*SUOX*	GTEx	Brain Hippocampus	0.94	NS	-	-
rs2271194	*SUOX*	GTEx	Brain Hypothalamus	0.97	NS	-	-
rs2271194	*SUOX*	GTEx	Brain Nucleus accumbens basal ganglia	0.93	NS	-	-
rs2271194	*SUOX*	GTEx	Brain Putamen basal ganglia	0.87	NS	-	-
rs2271194	*SUOX*	GTEx	Breast Mammary Tissue	0.92	NS	-	-
rs2271194	*SUOX*	GTEx	Cells Transformed fibroblasts	0.80	NS	-	-
rs2271194	*SUOX*	GTEx	Colon Sigmoid	0.95	0.73	-	-
rs2271194	*SUOX*	GTEx	Colon Transverse	0.97	NS	-	-
rs2271194	*SUOX*	GTEx	Esophagus Gastroesophageal Junction	0.93	NS	-	-
rs2271194	*SUOX*	GTEx	Esophagus Mucosa	0.81	NS	-	-
rs2271194	*SUOX*	GTEx	Esophagus Muscularis	0.81	NS	-	-
rs2271194	*SUOX*	GTEx	Heart Atrial Appendage	0.91	NS	-	-
rs2271194	*SUOX*	GTEx	Heart Left Ventricle	0.80	NS	-	-
rs2271194	*SUOX*	GTEx	Lung	0.87	NS	-	-
rs2271194	*SUOX*	GTEx	Muscle Skeletal	0.96	NS	-	-
rs2271194	*SUOX*	GTEx	Nerve Tibial	0.88	NS	-	-
rs2271194	*SUOX*	GTEx	Ovary	0.83	NS	-	-
rs2271194	*SUOX*	GTEx	Pancreas	0.86	NS	-	-
rs2271194	*SUOX*	GTEx	Pituitary	0.98	NS	-	-
rs2271194	*SUOX*	GTEx	Prostate	0.80	NS	-	-
rs2271194	*SUOX*	GTEx	Skin Not Sun Exposed Suprapubic	0.93	NS	-	-

PCOS SNP	Gene/Protein	QTL dataset	Tissue	HyPrColoc, 2 Mb, Main	HyPrColoc, 2 Mb, Without-23	HyPrColoc, independent regions, Main	HyPrColoc, independent regions, Without-23
rs2271194	*SUOX*	GTEx	Small Intestine Terminal Ileum	0.90	NS	-	-
rs2271194	*SUOX*	GTEx	Spleen	0.82	NS	-	-
rs2271194	*SUOX*	GTEx	Stomach	0.82	NS	-	-
rs2271194	*SUOX*	GTEx	Testis	0.93	NS	-	-
rs2271194	*SUOX*	GTEx	Thyroid	0.86	NS	-	-
rs2271194	*SUOX*	GTEx	Whole Blood	0.86	NS	-	-
rs2271194	*SUOX*	eQTLgen	Blood expression levels	NS	NS	-	-
rs2271194	*ZC3H10*	GTEx	Skin Sun Exposed Lower leg	NS	NS	-	-
rs7563201	*ZFP36L2*	eQTLgen	Blood expression levels	0.98	NS	-	-
rs7864171	*C9orf3*	GTEx	Heart Atrial Appendage	0.86	NS	-	-
rs804279	*C8orf49*	GTEx	Stomach	0.96	NS	-	-
rs804279	*NEIL2*	GTEx	Adrenal Gland	0.84	NS	-	-
rs804279	*NEIL2*	GTEx	Cells EBV-transformed lymphocytes	0.91	0.91	-	-
rs804279	*NEIL2*	GTEx	Cells Transformed fibroblasts	0.86	NS	-	-

Analyses performed using both the main PCOS dataset (with summary statistics from the maximum sample size for each SNP) and the Without-23 PCOS dataset (with summary statistics from approximately similar sample sizes for each SNP) as well as using different region sizes (400 kb, and, if full GWAS data available, also "independent regions" as per Berisa and Pickrell (2016)). "-" signifies that the analyses could not be performed due to missing data to perform the analysis. Table adapted (changed text, layout and included information) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

Appendix Table 4.7. Phenome-wide association data for rs11031005 for traits surpassing the multiple-testing threshold.

Trait	Estimate	P	N cases	N overall	Study ID	Author, year	PMID
Length of menstrual cycle	0.134	1.24E-42		32847	NEALE2_3710	UKB Neale v2	
Age at menopause (last menstrual period)	0.238	1.38E-15		111593	NEALE2_3581_raw	UKB Neale v2	
Heel bone mineral density	-0.018	5.7E-12		426824	GCST006979	Morris JA	PMID:30598549
Years since last cervical smear test	-0.025	1.12E-10		149196	NEALE2_2704	UKB Neale v2	
Bilateral oophorectomy (both ovaries removed)	-0.100	3.01E-09	15629	191515	NEALE2_2834	UKB Neale v2	
Had menopause	-0.038	3.14E-09	43829	163184	NEALE2_2724	UKB Neale v2	
Ever had hysterectomy (womb removed)	-0.098	1.92E-08	13973	171413	NEALE2_3591	UKB Neale v2	
Ever used hormone-replacement therapy (HRT)	-0.046	5.71E-08	76179	193606	NEALE2_2814	UKB Neale v2	
Age when periods started (menarche)	0.018	1.78E-07		188644	NEALE2_2714	UKB Neale v2	
Standing height	0.107	4.86E-07		360388	NEALE2_50_raw	UKB Neale v2	
Part of a multiple birth	-0.115	5.39E-07	7919	355467	NEALE2_1777	UKB Neale v2	
Forced vital capacity (FVC), Best measure	0.013	8.13E-07		272338	NEALE2_20151_raw	UKB Neale v2	
Number of operations, self-reported	-0.013	9.53E-07		361141	NEALE2_136	UKB Neale v2	

Bonferroni-significant (P<1.7E-6) phenome-wide association results from the Open Target Genetics platform at https://genetics.opentargets.org (Bonferroni, 1936; Carvalho-Silva *et al.*, 2019). Estimate; effect estimate in beta, or log(OR) for binary traits, according to the risk-increasing allele for PCOS (see Table 4.1); N cases: the number of cases for binary traits; N overall: total study sample size; Study ID: study ID from Open Targets (Carvalho-Silva *et al.*, 2019); Author, year: Auhor, and year if available, that the data is from; PMID: Pubmed ID that the data is from; UKB Neale v2; version two of the Neale lab analyses (Neale lab, 2020); SAIGE: analyses by the Lee lab (Zhou *et al.*, 2018). Table adapted (changed column names and table restricted to significant analyses) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

Appendix Table 4.8. Phenome-wide association data for rs227119 for traits surpassing the multiple-testing threshold.

Trait	Estimate	P	N cases	N overall	Study ID	Author, year	PMID
Eosinophill percentage	0.046	3.87E-25		349861	NEALE2_30210_raw	UKB Neale v2	
None of the above | Blood clot, DVT, bronchitis, emphysema, asthma, rhinitis, eczema, allergy diagnosed by doctor	-0.047	4.02E-20	116005	360527	NEALE2_6152_100	UKB Neale v2	
Hayfever, allergic rhinitis or eczema | Blood clot, DVT, bronchitis, emphysema, asthma, rhinitis, eczema, allergy diagnosed by doctor	0.047	2.91E-17	83407	360527	NEALE2_6152_9	UKB Neale v2	
asthma | Non-cancer illness code, self-reported	0.055	8.92E-14	41934	361141	NEALE2_20002_1111	UKB Neale v2	
Asthma | Blood clot, DVT, bronchitis, emphysema, asthma, rhinitis, eczema, allergy diagnosed by doctor	0.055	1.44E-13	41633	360527	NEALE2_6152_8	UKB Neale v2	
Eosinophill count	0.013	1.71E-13		349856	NEALE2_30150	UKB Neale v2	
Allergic disease (asthma, hay fever or eczema)	0.043	2.36E-13	180129	360838	GCST005038	Ferreira MA	PMID:29083406
Eosinophil percentage of white cells	0.026	2.66E-13		172378	GCST004600	Astle WJ	PMID:27863252
Impedance of arm (left)	0.660	1.63E-12		354807	NEALE2_23110_raw	UKB Neale v2	
Eosinophil counts	0.025	1.76E-12		172275	GCST004606	Astle WJ	PMID:27863252
Impedance of whole body	1.113	2.76E-12		354795	NEALE2_23106_raw	UKB Neale v2	
Arm predicted mass (left)	-0.007	2.94E-12		354653	NEALE2_23126_raw	UKB Neale v2	
None of the above | Qualifications	-0.042	5.83E-12	61093	357549	NEALE2_6138_100	UKB Neale v2	
Sum eosinophil basophil counts	0.024	1.38E-11		171771	GCST004624	Astle WJ	PMID:27863252
Arm fat-free mass (left)	-0.008	1.62E-11		354668	NEALE2_23125_raw	UKB Neale v2	
Impedance of arm (right)	0.607	2.65E-11		354792	NEALE2_23109_raw	UKB Neale v2	
Eosinophil percentage of granulocytes	0.023	7.16E-11		170536	GCST004617	Astle WJ	PMID:27863252
Whole body fat-free mass	-0.097	9.85E-11		354808	NEALE2_23101_raw	UKB Neale v2	
Whole body water mass	-0.071	1.04E-10		354834	NEALE2_23102_raw	UKB Neale v2	

Trait	Estimate	P	N cases	N overall	Study ID	Author, year	PMID
Basal metabolic rate	-12.873	1.31E-10		354825	NEALE2_23105_raw	UKB Neale v2	
Age first had sexual intercourse	0.061	1.42E-10		317694	NEALE2_2139_raw	UKB Neale v2	
Trunk fat-free mass	-0.049	1.48E-10		354530	NEALE2_23129_raw	UKB Neale v2	
Trunk predicted mass	-0.047	1.74E-10		354494	NEALE2_23130_raw	UKB Neale v2	
Neutrophil percentage of granulocytes	-0.023	3.24E-10		170672	GCST004623	Astle WJ	PMID:27863252
Arm fat-free mass (right)	-0.007	3.59E-10		354732	NEALE2_23121_raw	UKB Neale v2	
Arm predicted mass (right)	-0.006	3.89E-10		354726	NEALE2_23122_raw	UKB Neale v2	
hypothyroidism/myxoedema | Non-cancer illness code, self-reported	0.067	1.01E-09	17574	361141	NEALE2_20002_1226	UKB Neale v2	
College or University degree | Qualifications	0.030	4.1E-09	115981	357549	NEALE2_6138_1	UKB Neale v2	
Leg fat-free mass (right)	-0.017	7.45E-09		354798	NEALE2_23113_raw	UKB Neale v2	
Leg fat-free mass (left)	-0.017	8.33E-09		354771	NEALE2_23117_raw	UKB Neale v2	
Body mass index (BMI)	-0.065	8.47E-09		359983	NEALE2_21001_raw	UKB Neale v2	
Leg predicted mass (right)	-0.016	9.03E-09		354798	NEALE2_23114_raw	UKB Neale v2	
Leg predicted mass (left)	-0.015	1.1E-08		354766	NEALE2_23118_raw	UKB Neale v2	
Lymphocyte percentage	-0.099	1.39E-08		349861	NEALE2_30180_raw	UKB Neale v2	
Age completed full time education	0.013	1.45E-08		240547	NEALE2_845	UKB Neale v2	
Weight	-0.190	1.47E-08		360116	NEALE2_21002_raw	UKB Neale v2	
Average total household income before tax	0.016	2.71E-08		311028	NEALE2_738	UKB Neale v2	
Cognitive performance	0.016	2.89E-08		257841	GCST006572	Lee JJ	PMID:30038396
Body mass index (BMI)	-0.062	3.6E-08		354831	NEALE2_23104_raw	UKB Neale v2	
Age hay fever, rhinitis or eczema diagnosed	-0.445	4.41E-08		72232	NEALE2_3761_raw	UKB Neale v2	
Weight	-0.183	5.09E-08		354838	NEALE2_23098_raw	UKB Neale v2	
Asthma	0.051	6.2E-08	26332	401837	SAIGE_495	UKB SAIGE	
Impedance of leg (left)	0.417	6.6E-08		354811	NEALE2_23108_raw	UKB Neale v2	

Trait	Estimate	P	N cases	N overall	Study ID	Author, year	PMID
Hypothyroidism NOS	0.069	6.83E-08	14171	405600	SAIGE_244_4	UKB SAIGE	
Impedance of leg (right)	0.413	8.89E-08		354817	NEALE2_23107_raw	UKB Neale v2	
Hypothyroidism	0.064	2.72E-07	14871	406300	SAIGE_244	UKB SAIGE	
Never | Smoking status	0.024	2.88E-07	164638	359706	NEALE2_20116_0	UKB Neale v2	
hayfever/allergic rhinitis | Non-cancer illness code, self-reported	0.052	3.31E-07	20667	361141	NEALE2_20002_1387	UKB Neale v2	
Ankle spacing width	-0.063	3.4E-07		206589	NEALE2_3143_raw	UKB Neale v2	
levothyroxine sodium | Treatment/medication code	0.060	4.3E-07	14689	361141	NEALE2_20003_1141191044	UKB Neale v2	
A levels/AS levels or equivalent | Qualifications	0.026	5.1E-07	98304	357549	NEALE2_6138_2	UKB Neale v2	
Time spent watching television (TV)	-0.009	5.42E-07		341859	NEALE2_1070	UKB Neale v2	
Waist circumference	-0.136	1.55E-06		360564	NEALE2_48_raw	UKB Neale v2	

Bonferroni-significant (P<1.7E-6) phenome-wide association results from the Open Target Genetics platform at https://genetics.opentargets.org (Bonferroni, 1936; Carvalho-Silva *et al.*, 2019). Estimate; effect estimate in beta, or log(OR) for binary traits, according to the risk-increasing allele for PCOS (see Table 4.1); N cases: the number of cases for binary traits; N overall: total study sample size; Study ID: study ID from Open Targets (Carvalho-Silva *et al.*, 2019); Author, year: Auhor, and year if available, that the data is from; PMID: Pubmed ID that the data is from; UKB Neale v2; version two of the Neale lab analyses (Neale lab, 2020); SAIGE: analyses by the Lee lab (Zhou *et al.*, 2018). Table adapted (changed column names and table restricted to significant analyses) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

Appendix Table 4.9. Phenome-wide association data for rs7563201 for traits surpassing the multiple-testing threshold.

Trait	Estimate	P	N cases	N overall	Study ID	Author, year	PMID
Neutrophill percentage	0.132	6.46E-11		349861	NEALE2_30200_raw	UKB Neale v2	
Lymphocyte percentage	-0.111	1.78E-10		349861	NEALE2_30180_raw	UKB Neale v2	
Neutrophil percentage of white cells	0.019	2.01E-07		171542	GCST004633	Astle WJ	PMID:27863252
Lymphocyte percentage of white cells	-0.018	4.01E-07		171748	GCST004632	Astle WJ	PMID:27863252
Total body bone mineral density (age 0-15)	0.067	4.77E-07		11807	GCST005345	Medina-Gomez C	PMID:29304378
Heel bone mineral density	0.010	7E-07		426824	GCST006979	Morris JA	PMID:30598549
Neutrophill count	0.016	1.17E-06		349856	NEALE2_30140_raw	UKB Neale v2	

Bonferroni-significant (P<1.7E-6) phenome-wide association results from the Open Target Genetics platform at https://genetics.opentargets.org (Bonferroni, 1936; Carvalho-Silva *et al.*, 2019). Estimate; effect estimate in beta, or log(OR) for binary traits, according to the risk-increasing allele for PCOS (see Table 4.1); N cases: the number of cases for binary traits; N overall: total study sample size; Study ID: study ID from Open Targets (Carvalho-Silva *et al.*, 2019); Author, year: Auhor, and year if available, that the data is from; PMID: Pubmed ID that the data is from; UKB Neale v2; version two of the Neale lab analyses (Neale lab, 2020); SAIGE: analyses by the Lee lab (Zhou *et al.*, 2018). Table adapted (changed column names and table restricted to significant analyses) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

Appendix Table 4.10. Phenome-wide association data for rs13164856 for traits surpassing the multiple-testing threshold.

Trait	Estimate	P	N cases	N overall	Study ID	Author, year	PMID
Eosinophill count	-0.029	1.64E-51		349856	NEALE2_30150	UKB Neale v2	
Platelet count	-1.789	2.91E-31		350474	NEALE2_30080_raw	UKB Neale v2	
Sum eosinophil basophil counts	-0.043	2.74E-27		171771	GCST004624	Astle WJ	PMID:27863252
Standing height	0.173	9.3E-26		360388	NEALE2_50_raw	UKB Neale v2	
Eosinophill percentage	-0.051	1.79E-25		349861	NEALE2_30210_raw	UKB Neale v2	
Eosinophil counts	-0.041	8.78E-25		172275	GCST004606	Astle WJ	PMID:27863252
Eosinophil percentage of white cells	-0.035	1.92E-18		172378	GCST004600	Astle WJ	PMID:27863252
Platelet crit	-0.001	8.92E-18		350471	NEALE2_30090_raw	UKB Neale v2	
Neutrophil percentage of granulocytes	0.032	1.31E-15		170672	GCST004623	Astle WJ	PMID:27863252
Mean platelet (thrombocyte) volume	0.023	1.54E-15		350470	NEALE2_30100_raw	UKB Neale v2	
Eosinophil percentage of granulocytes	-0.032	1.77E-15		170536	GCST004617	Astle WJ	PMID:27863252
Platelet count	-0.032	7.3E-15		166066	GCST004603	Astle WJ	PMID:27863252
asthma | Non-cancer illness code, self-reported	-0.061	9.66E-14	41934	361141	NEALE2_20002_1111	UKB Neale v2	
Asthma | Blood clot, DVT, bronchitis, emphysema, asthma, rhinitis, eczema, allergy diagnosed by doctor	-0.059	4.17E-13	41633	360527	NEALE2_6152_8	UKB Neale v2	
Whole body water mass	0.087	7.66E-13		354834	NEALE2_23102_raw	UKB Neale v2	
Trunk fat-free mass	0.061	7.97E-13		354530	NEALE2_23129_raw	UKB Neale v2	
Neutrophill count	-0.027	9.97E-13		349856	NEALE2_30140_raw	UKB Neale v2	
Trunk predicted mass	0.058	1.02E-12		354494	NEALE2_23130_raw	UKB Neale v2	

Trait	Estimate	P	N cases	N overall	Study ID	Author, year	PMID
Whole body fat-free mass	0.117	1.12E-12		354808	NEALE2_23101_raw	UKB Neale v2	
White blood cell (leukocyte) count	-0.038	2.54E-12		350470	NEALE2_30000_raw	UKB Neale v2	
Sitting height	0.065	1.4E-11		360066	NEALE2_20015_raw	UKB Neale v2	
Comparative height size at age 10	0.012	1.41E-11		355331	NEALE2_1697	UKB Neale v2	
Forced expiratory volume in 1-second (FEV1), predicted	0.007	1.51E-11		117241	NEALE2_20153_raw	UKB Neale v2	
Basal metabolic rate	14.789	1.93E-11		354825	NEALE2_23105_raw	UKB Neale v2	
Leg predicted mass (right)	0.019	5.73E-11		354798	NEALE2_23114_raw	UKB Neale v2	
Leg fat-free mass (right)	0.021	6.85E-11		354798	NEALE2_23113_raw	UKB Neale v2	
Leg predicted mass (left)	0.019	9.41E-11		354766	NEALE2_23118_raw	UKB Neale v2	
Leg fat-free mass (left)	0.021	1.05E-10		354771	NEALE2_23117_raw	UKB Neale v2	
None of the above | Blood clot, DVT, bronchitis, emphysema, asthma, rhinitis, eczema, allergy diagnosed by doctor	0.035	3.1E-10	116005	360527	NEALE2_6152_100	UKB Neale v2	
Ankle spacing width	0.085	4.4E-10		206589	NEALE2_3143_raw	UKB Neale v2	
Arm fat-free mass (right)	0.007	6.1E-10		354732	NEALE2_23121_raw	UKB Neale v2	
Arm predicted mass (right)	0.007	8.39E-10		354726	NEALE2_23122_raw	UKB Neale v2	
Arm fat-free mass (left)	0.008	9.72E-10		354668	NEALE2_23125_raw	UKB Neale v2	
Myeloid white cell count	-0.025	9.95E-10		169219	GCST004626	Astle WJ	PMID:27863252

Trait	Estimate	P	N cases	N overall	Study ID	Author, year	PMID
Arm predicted mass (left)	0.007	1.38E-09		354653	NEALE2_23126_raw	UKB Neale v2	
Mean platelet volume	0.024	3.69E-09		164454	GCST004599	Astle WJ	PMID:27863252
Sum neutrophil eosinophil counts	-0.023	6.1E-09		170384	GCST004613	Astle WJ	PMID:27863252
Granulocyte count	-0.023	6.4E-09		169822	GCST004614	Astle WJ	PMID:27863252
Plateletcrit	-0.024	6.86E-09		164339	GCST004607	Astle WJ	PMID:27863252
Hand grip strength (left)	0.111	1.76E-08		359704	NEALE2_46_raw	UKB Neale v2	
White blood cell count	-0.023	1.94E-08		172435	GCST004610	Astle WJ	PMID:27863252
Crohn's disease [EA]	0.076	2.38E-08	5956	20883	GCST003044	Liu JZ	PMID:26192919
Hand grip strength (right)	0.109	3.55E-08		359729	NEALE2_47_raw	UKB Neale v2	
Crohn's disease	0.100	4.67E-08	12194	40266	GCST004132	de Lange KM	PMID:28067908
Hayfever, allergic rhinitis or eczema | Blood clot, DVT, bronchitis, emphysema, asthma, rhinitis, eczema, allergy diagnosed by doctor	-0.033	9.83E-08	83407	360527	NEALE2_6152_9	UKB Neale v2	
Monocyte count	-0.003	1.11E-07		349856	NEALE2_30130_raw	UKB Neale v2	
Weight	0.194	1.68E-07		354838	NEALE2_23098_raw	UKB Neale v2	
Weight	0.188	3.22E-07		360116	NEALE2_21002_raw	UKB Neale v2	
Asthma	-0.074	4.36E-07	19954	127669	GCST006862	Demenais F	PMID:29273806
Ankle spacing width (right)	0.092	5.78E-07		114614	NEALE2_4119_raw	UKB Neale v2	
Sum basophil neutrophil counts	-0.019	1.38E-06		170143	GCST004620	Astle WJ	PMID:27863252
Neutrophil count	-0.019	1.52E-06		170702	GCST004629	Astle WJ	PMID:27863252

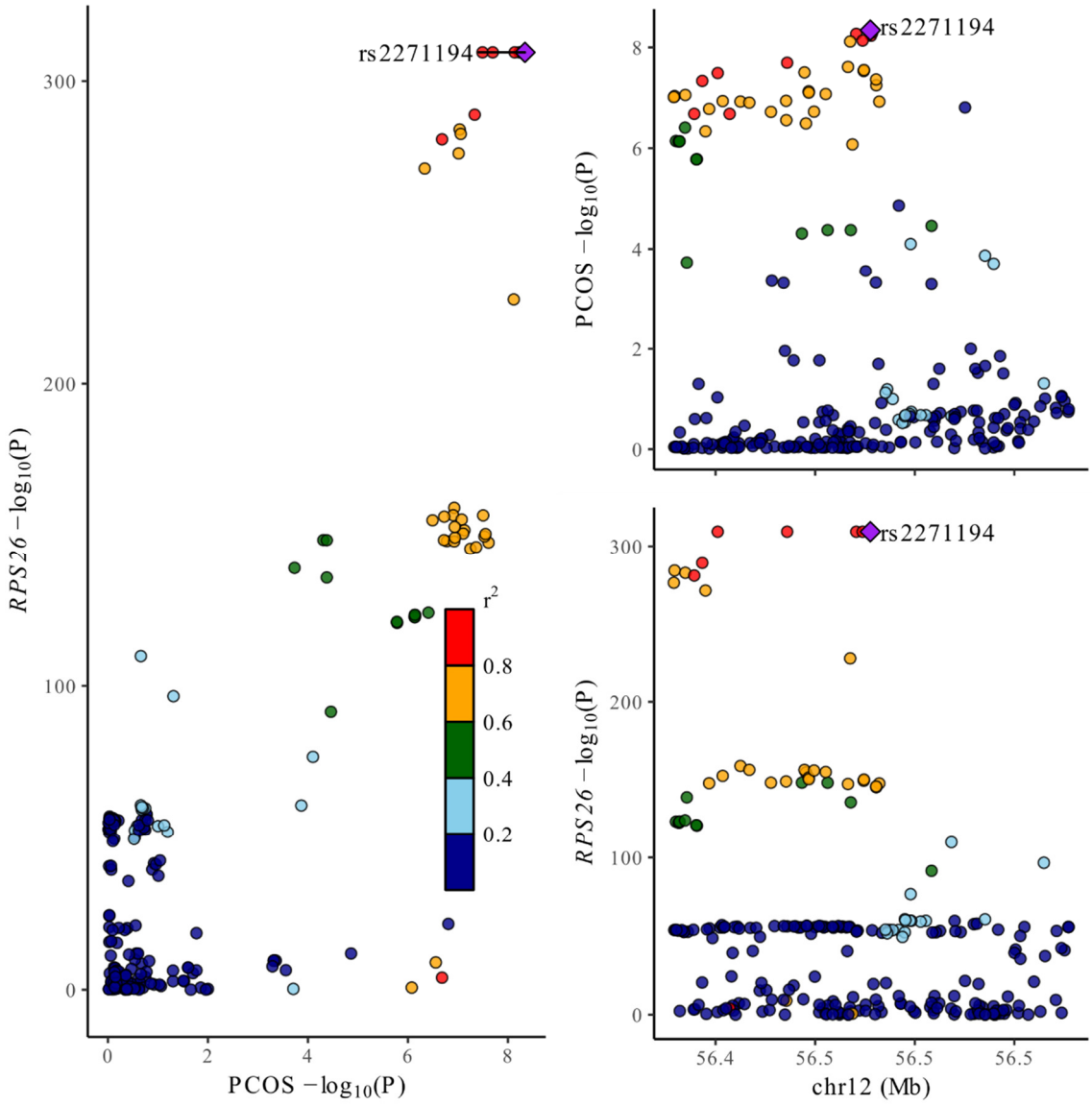
Trait	Estimate	P	N cases	N overall	Study ID	Author, year	PMID
None of the above | Vascular/heart problems diagnosed by doctor	0.026	1.62E-06	106855	360420	NEALE2_6150_100	UKB Neale v2	

Bonferroni-significant (P<1.7E-6) phenome-wide association results from the Open Target Genetics platform at https://genetics.opentargets.org (Bonferroni, 1936; Carvalho-Silva *et al.*, 2019). Estimate; effect estimate in beta, or log(OR) for binary traits, according to the risk-increasing allele for PCOS (see Table 4.1); N cases: the number of cases for binary traits; N overall: total study sample size; Study ID: study ID from Open Targets (Carvalho-Silva *et al.*, 2019); Author, year: Auhor, and year if available, that the data is from; PMID: Pubmed ID that the data is from; UKB Neale v2; version two of the Neale lab analyses (Neale lab, 2020); SAIGE: analyses by the Lee lab (Zhou *et al.*, 2018). Table adapted (changed column names and table restricted to significant analyses) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).

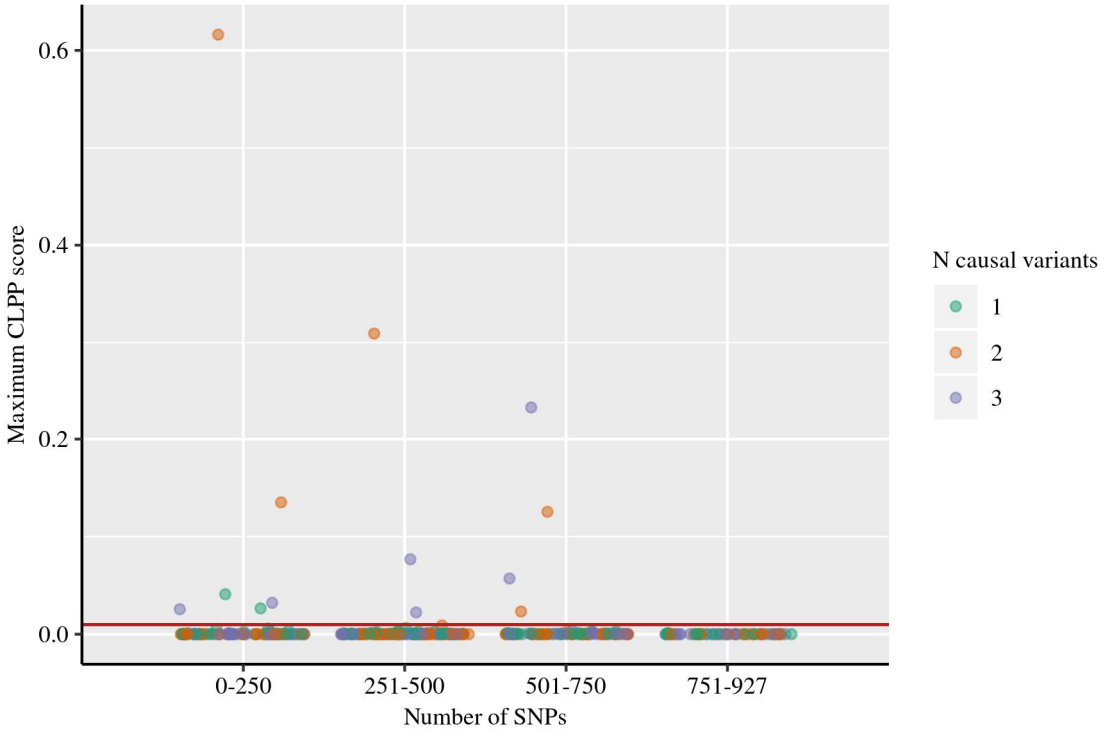
Appendix Table 4.11. Phenome-wide association data for rs7864171 for traits surpassing the multiple-testing threshold.

Trait	Estimate	P	N cases	N overall	Study ID	Author, year	PMID
Atrial fibrillation	-0.083	5.96E-34	60620	1030836	GCST006414	Nielsen JB	PMID:30061737
Standing height	-0.129	1.41E-17		360388	NEALE2_50_raw	UKB Neale v2	
Atrial fibrillation and flutter	-0.094	7.49E-14	14820	395739	SAIGE_427_2	UKB SAIGE	
Trunk fat-free mass	-0.051	4.27E-11		354530	NEALE2_23129_raw	UKB Neale v2	
Trunk predicted mass	-0.049	6.37E-11		354494	NEALE2_23130_raw	UKB Neale v2	
atrial fibrillation | Non-cancer illness code, self-reported	-0.172	2.52E-10	2828	361141	NEALE2_20002_1471	UKB Neale v2	
Whole body fat-free mass	-0.089	4.28E-09		354808	NEALE2_23101_raw	UKB Neale v2	
Whole body water mass	-0.065	4.66E-09		354834	NEALE2_23102_raw	UKB Neale v2	
Sitting height	-0.050	1.06E-08		360066	NEALE2_20015_raw	UKB Neale v2	
Cardiac dysrhythmias	-0.055	2E-08	24681	405600	SAIGE_427	UKB SAIGE	
Arm fat-free mass (right)	-0.006	5.62E-08		354732	NEALE2_23121_raw	UKB Neale v2	
Arm predicted mass (right)	-0.005	9.54E-08		354726	NEALE2_23122_raw	UKB Neale v2	
Arm predicted mass (left)	-0.006	1.04E-07		354653	NEALE2_23126_raw	UKB Neale v2	
Basal metabolic rate	-10.507	1.97E-07		354825	NEALE2_23105_raw	UKB Neale v2	
Forced expiratory volume in 1-second (FEV1), predicted	-0.005	2.14E-07		117241	NEALE2_20153_raw	UKB Neale v2	
Arm fat-free mass (left)	-0.006	2.29E-07		354668	NEALE2_23125_raw	UKB Neale v2	
Forced vital capacity (FVC), Best measure	-0.009	6.03E-07		272338	NEALE2_20151_raw	UKB Neale v2	
Snoring	-0.024	1.39E-06	125656	336320	NEALE2_1210	UKB Neale v2	

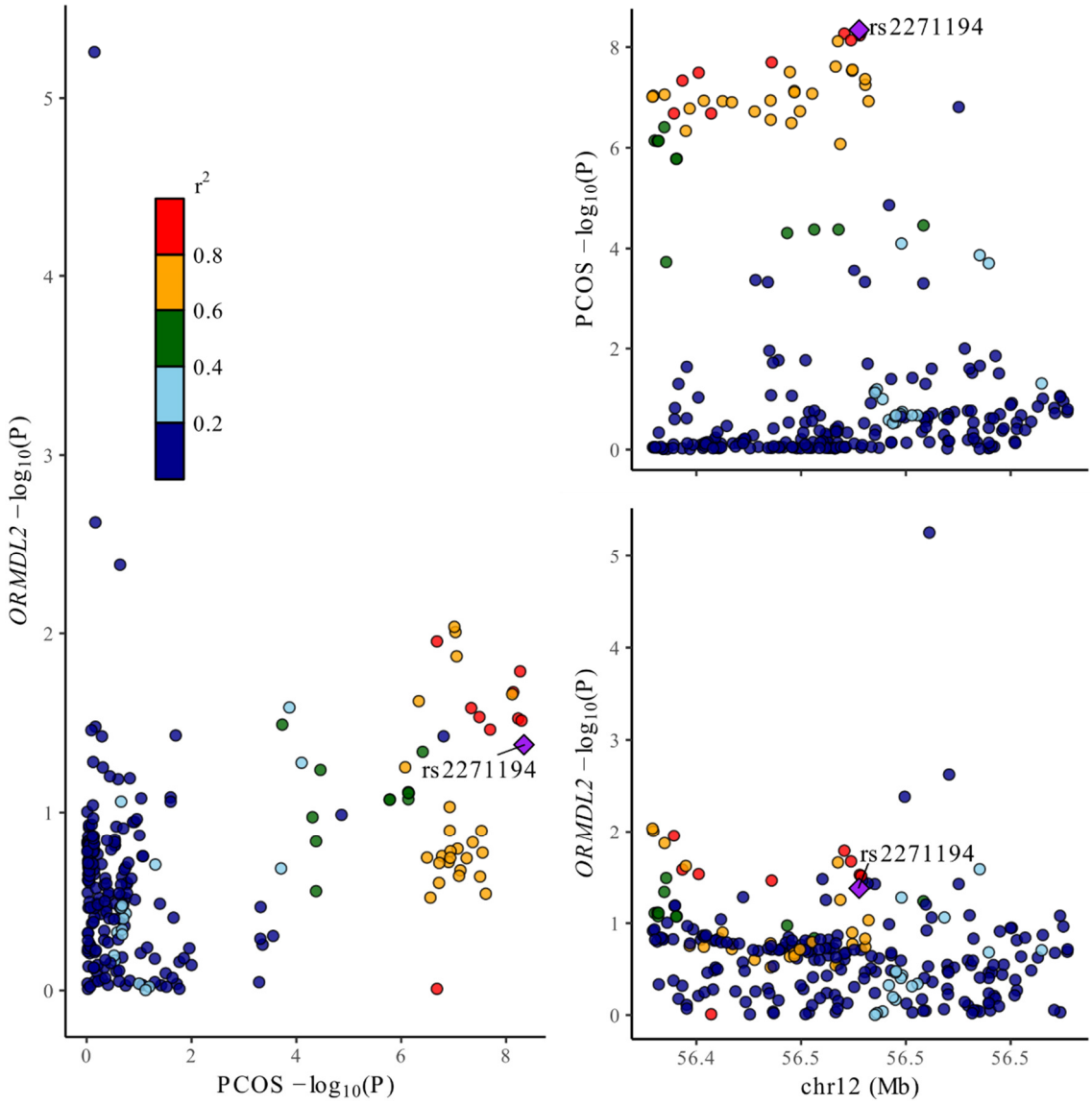
Bonferroni-significant (P<1.7E-6) phenome-wide association results from the Open Target Genetics platform at https://genetics.opentargets.org (Bonferroni, 1936; Carvalho-Silva *et al.*, 2019). Estimate; effect estimate in beta, or log(OR) for binary traits, according to the risk-increasing allele for PCOS (see Table 4.1); N cases: the number of cases for binary traits; N overall: total study sample size; Study ID: study ID from Open Targets (Carvalho-Silva *et al.*, 2019); Author, year: Auhor, and year if available, that the data is from; PMID: Pubmed ID that the data is from; UKB Neale v2; version two of the Neale lab analyses (Neale lab, 2020); SAIGE: analyses by the Lee lab (Zhou *et al.*, 2018). Table adapted (changed column names and table restricted to significant analyses) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).



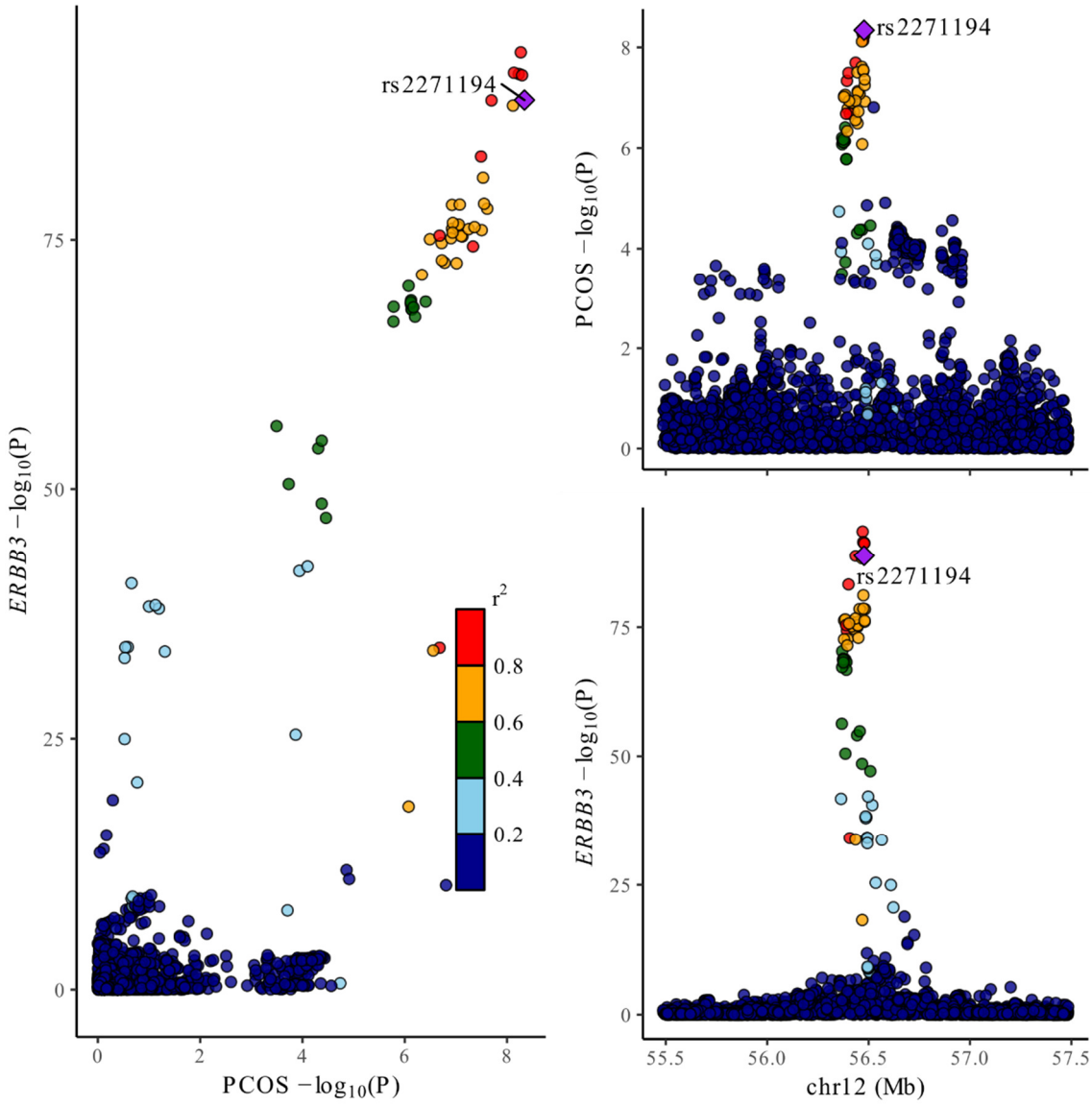
Appendix Figure 4.1. The SNP associations for PCOS risk (using the main PCOS dataset) and *RPS26* expression in blood (eQTLgen) at the rs2271194 risk locus. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout and region size) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).



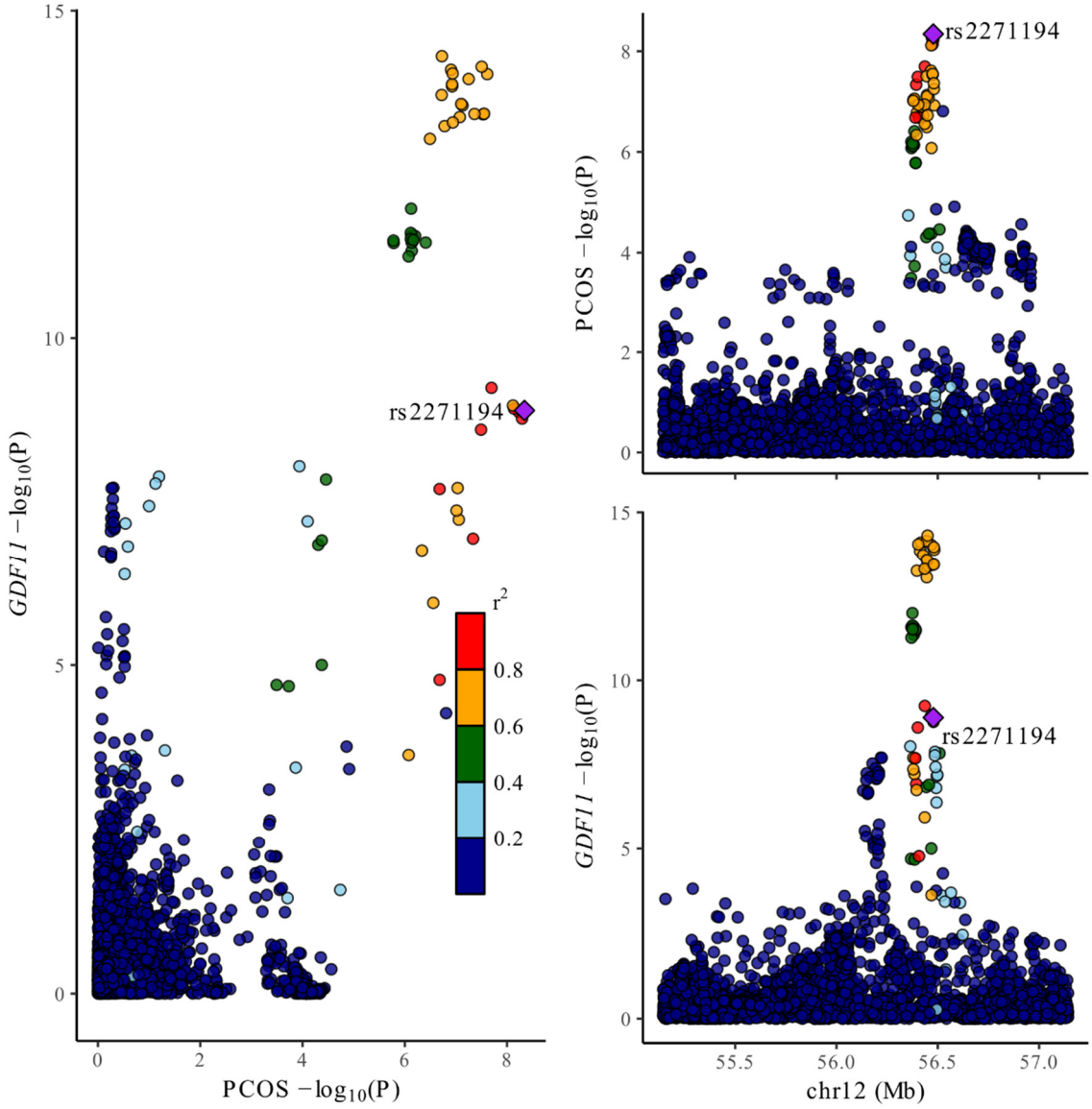
Appendix Figure 4.2. eCAVIAR colocalisation results using the PCOS dataset where the 23andMe cohort had been excluded and gene expression and protein level data from GTEx and INTERVAL, respectively. Number of SNPs is the number of SNPs in the investigated region (+/- 100 kb from the top PCOS SNP) and maximum CLPP score the highest CLPP score of any analysed SNP in the investigated region.



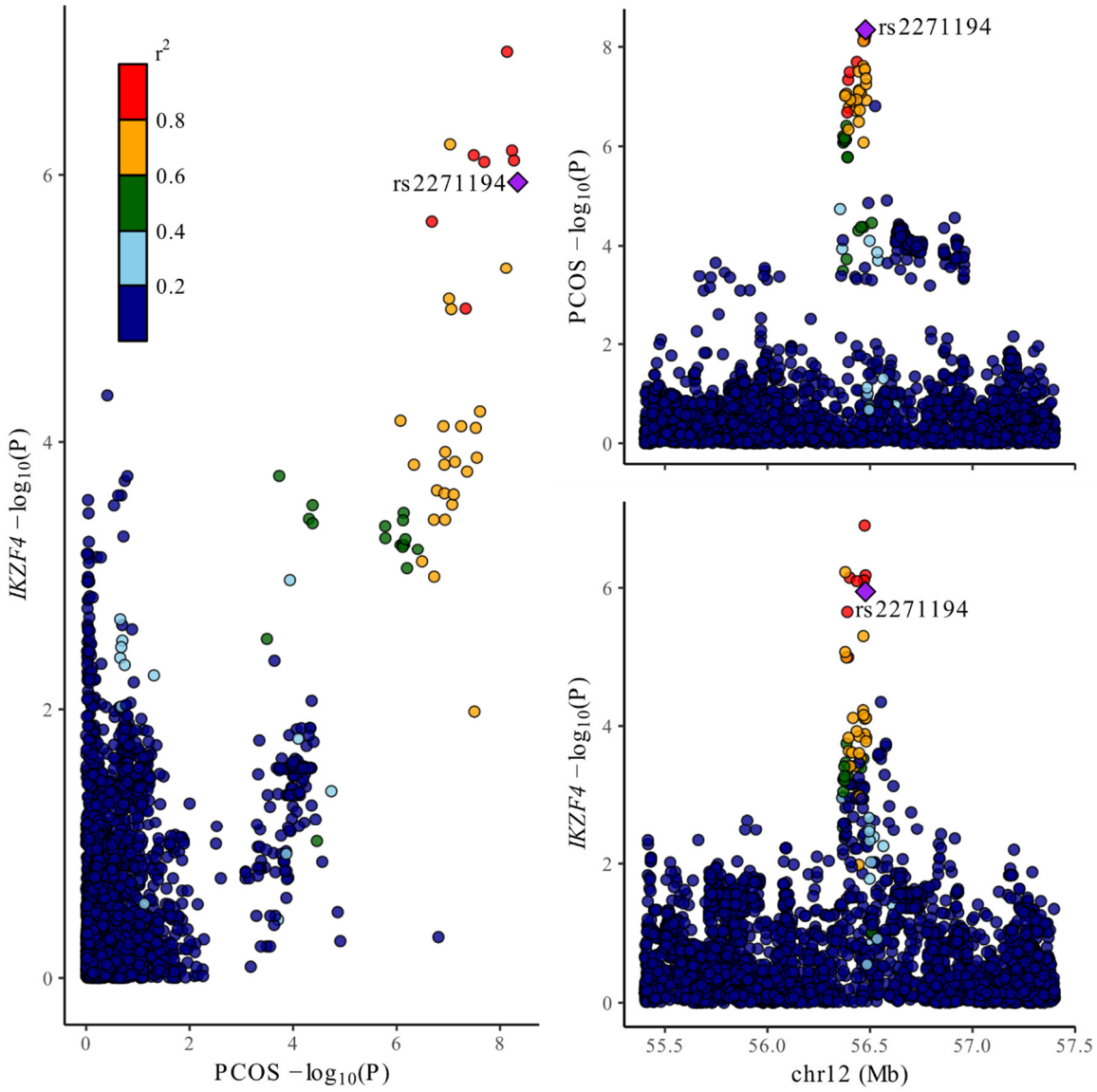
Appendix Figure 4.3. The SNP associations for PCOS risk (using the main PCOS dataset) and *ORMDL2* expression in blood (eQTLgen) at the rs2271194 risk locus. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue.



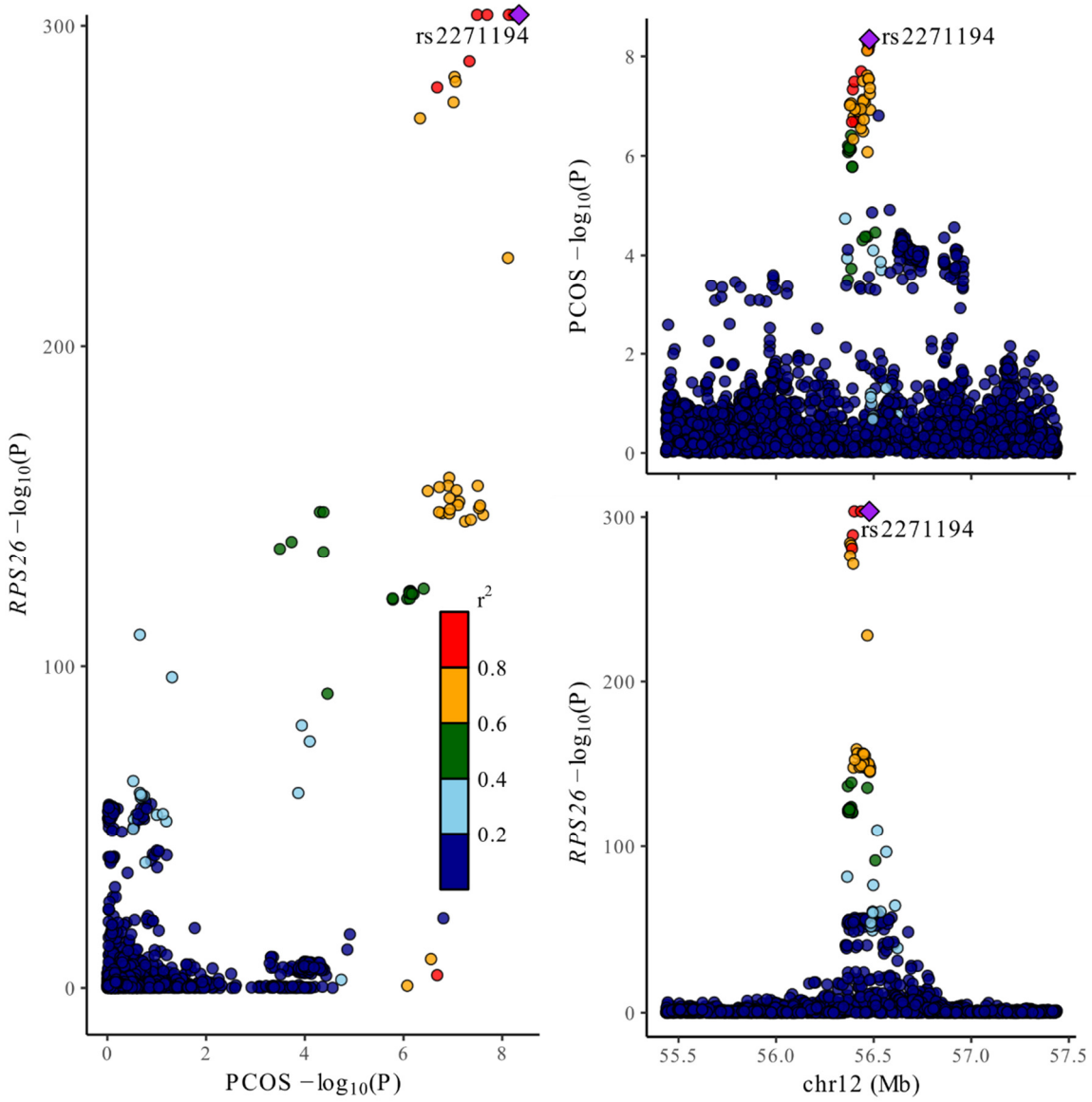
Appendix Figure 4.4. The SNP associations for PCOS risk (using the main PCOS dataset) and *ERBB3* expression in blood (eQTLgen) at the rs2271194 risk locus. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).



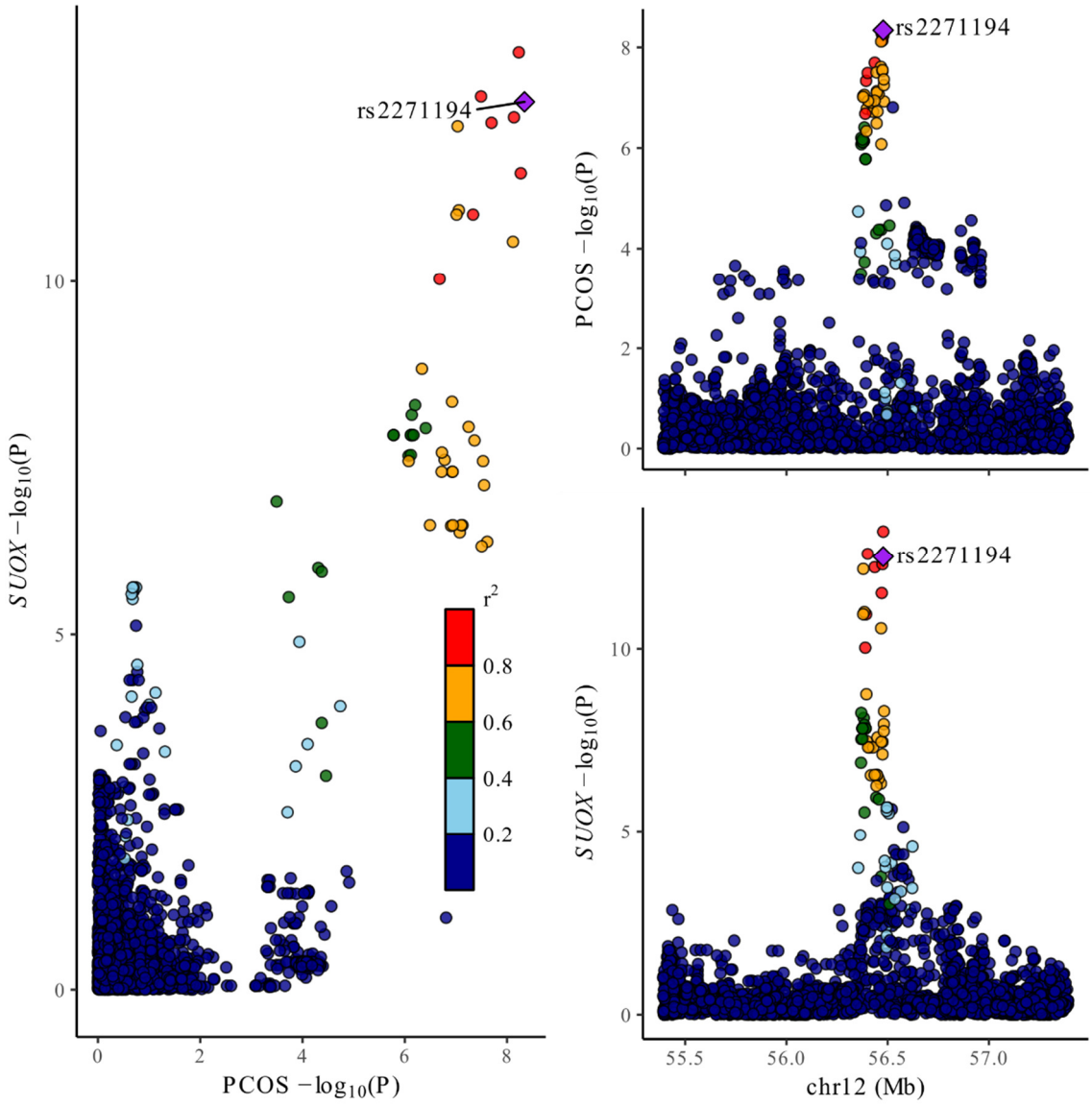
Appendix Figure 4.5. The SNP associations for PCOS risk (using the main PCOS dataset) and *GDF11* expression in blood (eQTLgen) at the rs2271194 risk locus. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).



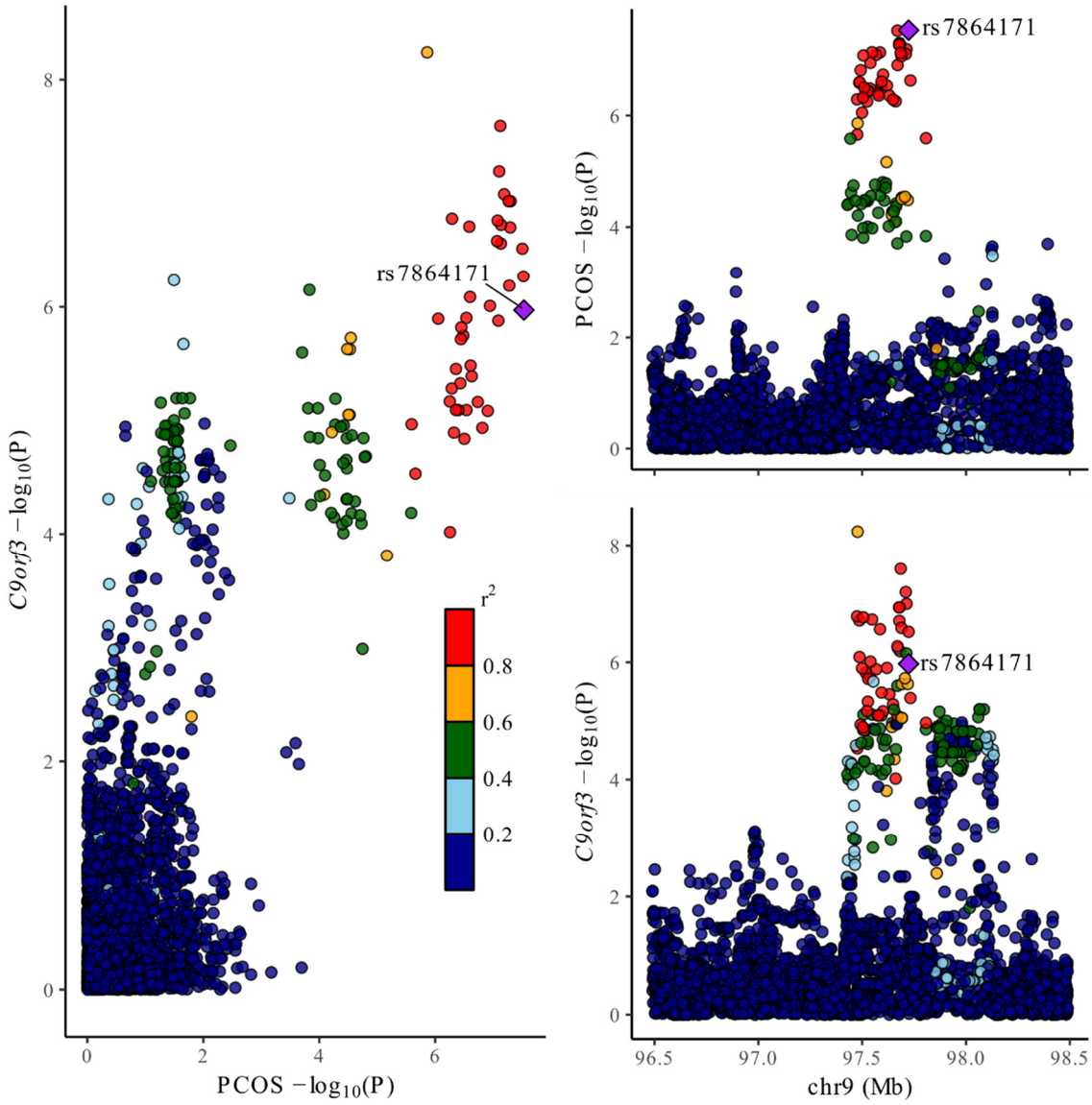
Appendix Figure 4.6. The SNP associations for PCOS risk (using the main PCOS dataset) and *IKZF4* expression in oesophagus mucosa (GTEx) at the rs2271194 risk locus. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).



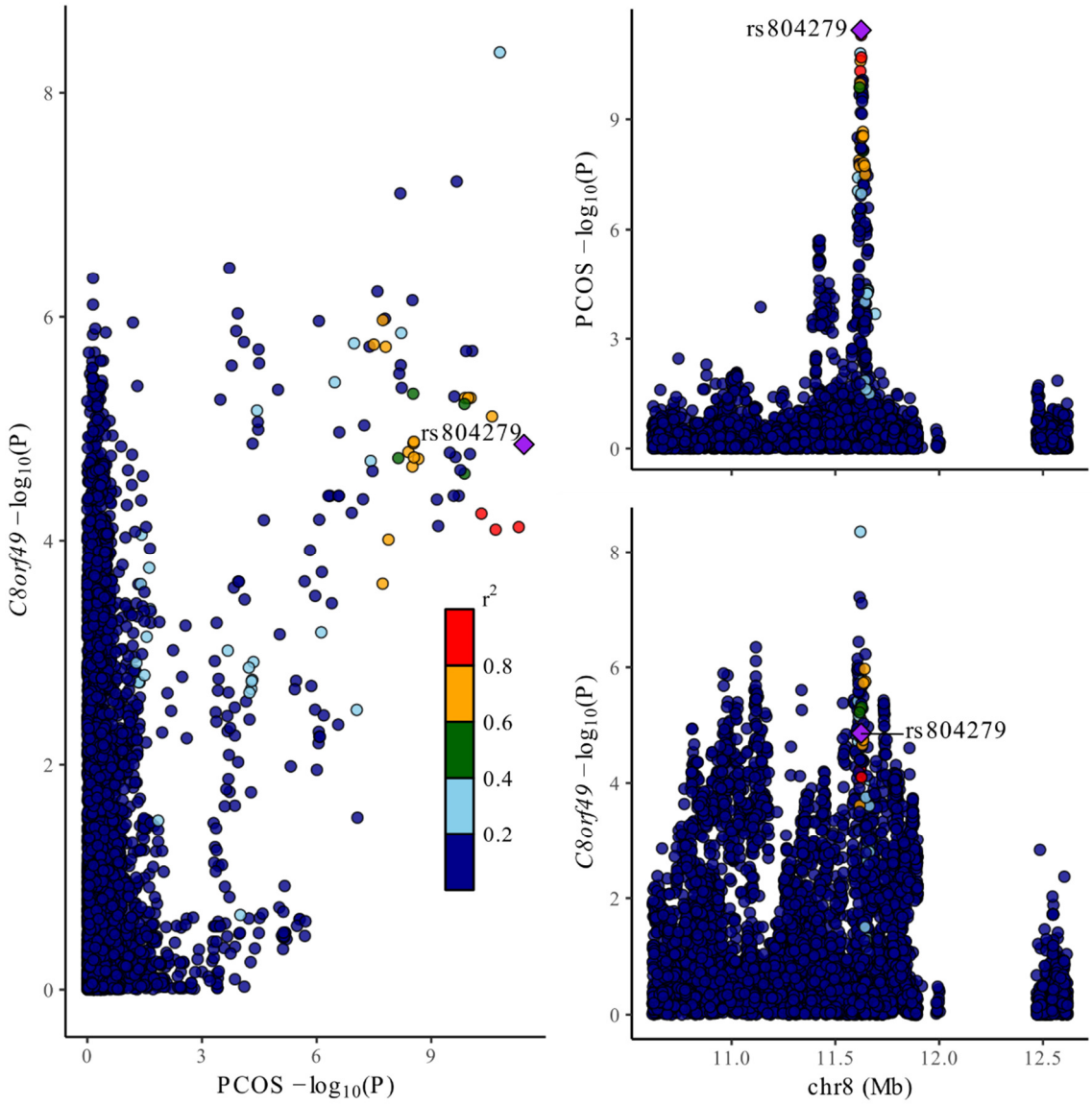
Appendix Figure 4.7. The SNP associations for PCOS risk (using the main PCOS dataset) and *RPS26* expression in blood (eQTLgen) at the rs2271194 risk locus. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).



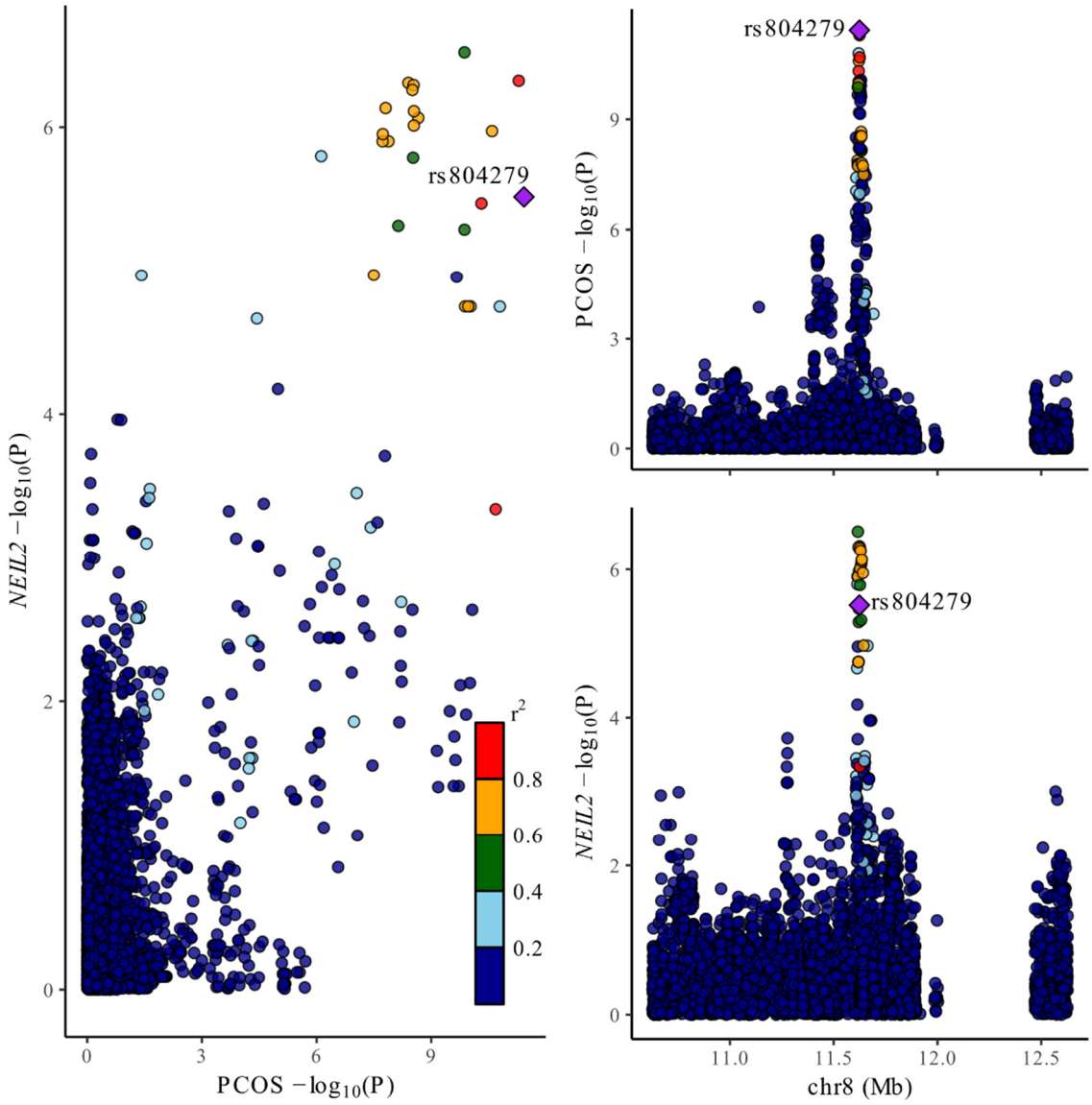
Appendix Figure 4.8. The SNP associations for PCOS risk (using the main PCOS dataset) and *SUOX* expression in pituitary gland (GTEx) at the rs2271194 risk locus. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).



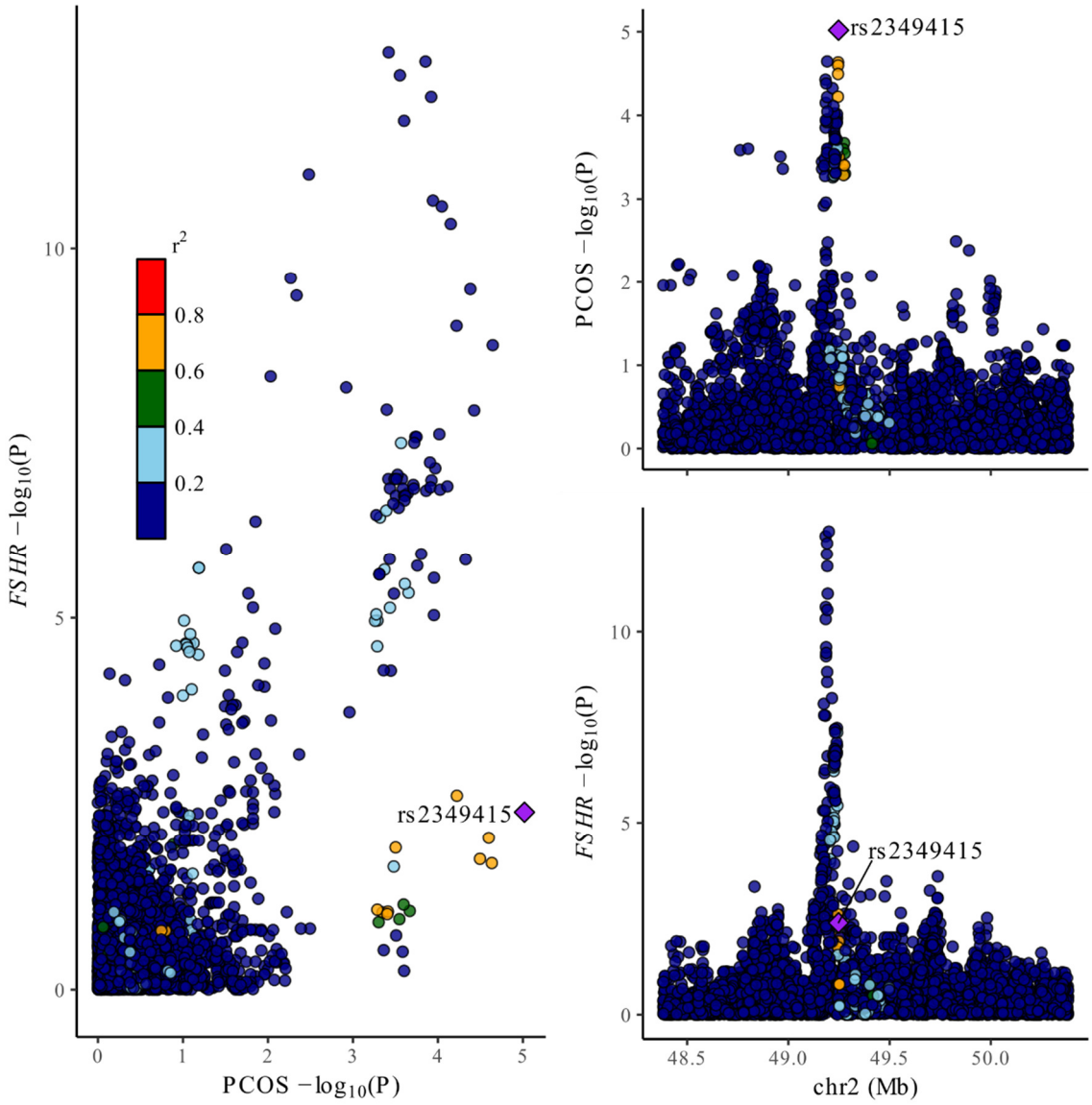
Appendix Figure 4.9. The SNP associations for PCOS risk (using the main PCOS dataset) and *C9orf3* expression in the atrial appendage of the heart (GTEx) at the rs7864171 risk locus. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).



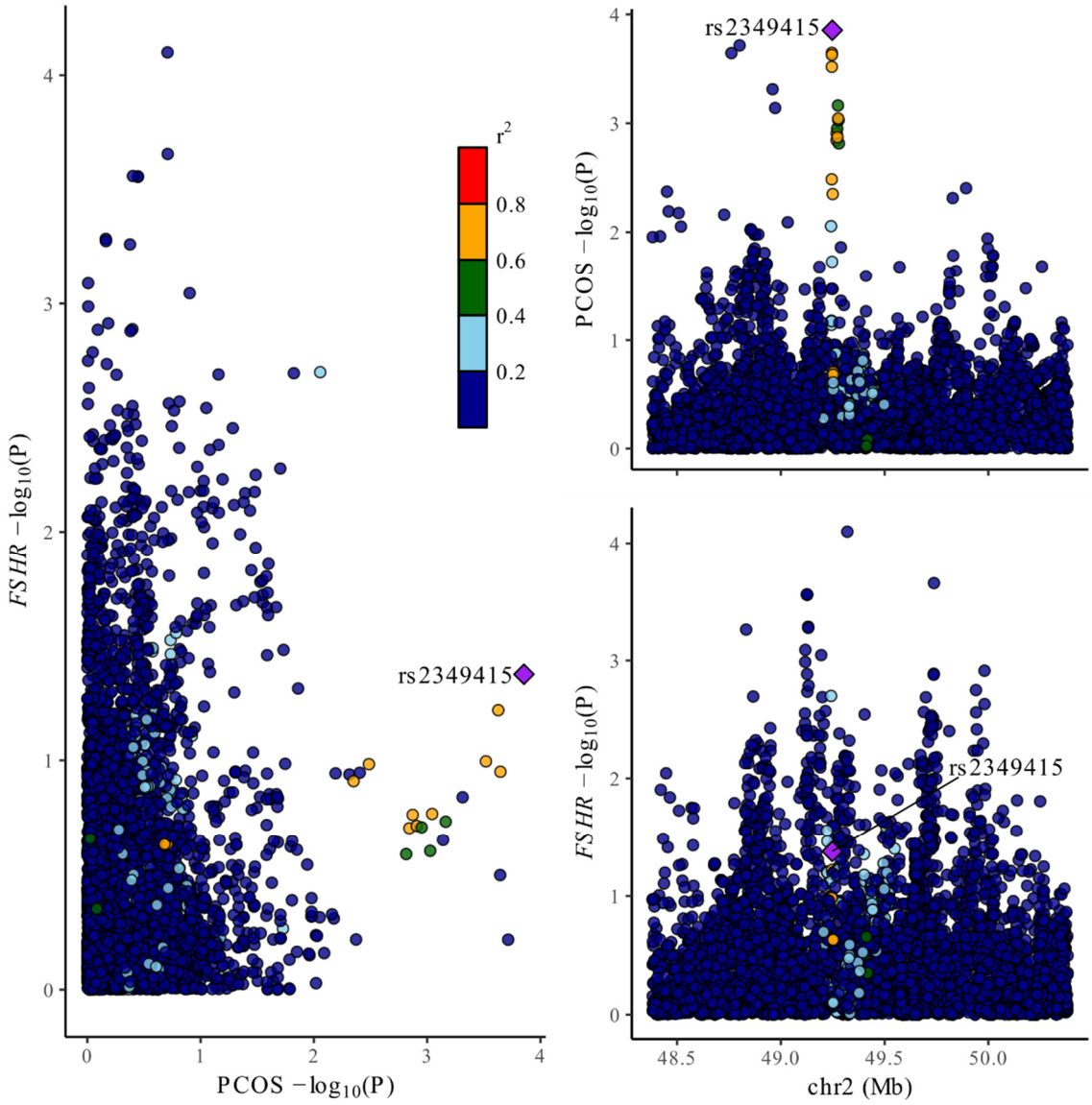
Appendix Figure 4.10. The SNP associations for PCOS risk (using the main PCOS dataset) and *C8orf49* expression in stomach (GTEx) at the rs804279 risk locus. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).



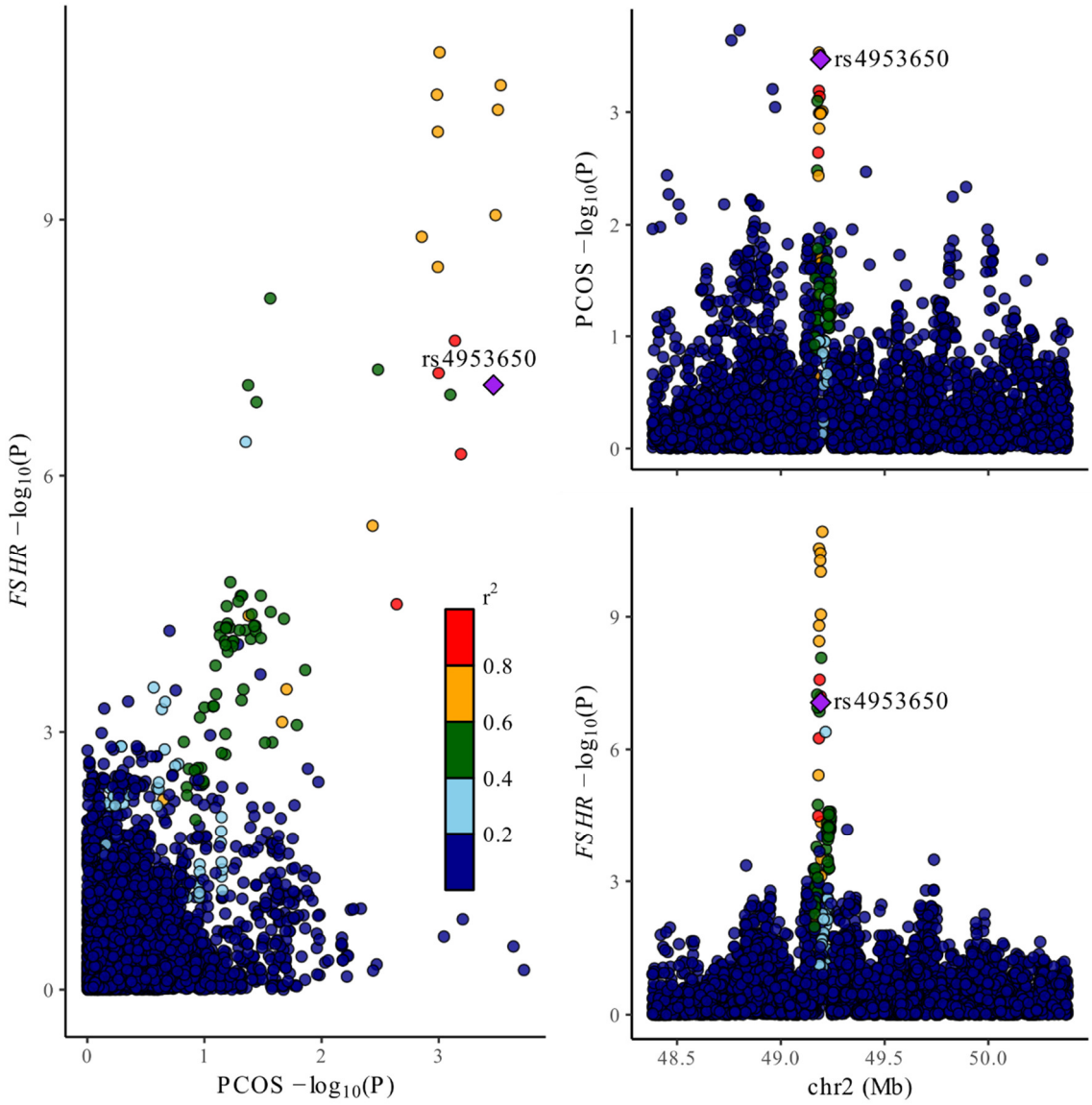
Appendix Figure 4.11. The SNP associations for PCOS risk (using the main PCOS dataset) and *NEIL2* expression in EBV-transformed lymphocytes (GTEx) at the rs804279 risk locus. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).



Appendix Figure 4.12. The SNP associations for PCOS risk (using the main PCOS dataset) and *FSHR* expression in testis (GTEx). The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).



Appendix Figure 4.13. The SNP associations for PCOS risk (using the main PCOS dataset) and *FSHR* expression in testis (GTEx), using effect estimates conditioned on rs4953650. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).



Appendix Figure 4.14. The SNP associations for PCOS risk (using the main PCOS dataset) and *FSHR* expression in testis (GTEx), using effect estimates conditioned on rs2349415. The SNP most robustly associated with PCOS is marked in each plot, with other SNPs colour-coded by LD (r2 computed in Europeans) with that SNP. SNPs with missing LD information are coded dark blue. Figure adapted (changed layout) from Censin *et al.* (2021) and the corresponding pre-print (Censin *et al.*, 2020).